Audio Engineering Society

Convention Paper

Presented at the 110th Convention
2001 May 12–15 Amsterdam, The Netherlands

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.

Determining Optimum Room Dimensions for Critical Listening Environments: A New Methodology

Trevor J. Cox
School of Acoustics and Electronic Engineering, University of Salford, Salford, M5 4WT,UK. t.j.cox@salford.ac.uk.

and

Peter D'Antonio
RPG Diffusor Systems Inc., 651-c Commerce Drive, Upper Marlboro, USA pdantonio@rpginc.com

ABSTRACT

Modes in small rooms may lead to uneven frequency responses and extended sound decays at low frequencies. In critical listening environments this often causes unwanted coloration effects, which can be detrimental to the sound quality. Choosing an appropriately proportioned room may reduce the audible effects of modes. This paper details a new methodology for determining the room dimensions for small critical listening spaces. It is based on numerical optimisation of the room dimensions to achieve the flattest possible frequency response. The method is contrasted with previous techniques.

INTRODUCTION

Modes in small rooms often lead to extended sound decays and uneven frequency responses. In critical listening spaces, this causes unwanted coloration effects that can be detrimental to the sound quality. The problem arises at low frequencies because of the relatively low modal density. Many designers try to overcome the problems of modes by choosing an appropriately proportioned room and by the use of bass absorbers. This paper is interested in the former, the choice of room dimensions to minimise the coloration effects of modes. The paper starts by discussing previous studies by others, which have suggested optimum room ratios or design methodologies. Then a new method is outlined - this is based on numerical optimisation - and the old and new methods are compared philosophically. Results in the form of modal responses are given to demonstrate the power of the new method.

PREVIOUS WORK

Many methods and optimum room ratios have been suggested over the years to minimise coloration. Essentially these methods try to avoid degenerate modes, where multiple modal frequencies fall within a small bandwidth, and also bandwidths with absences of modes. The assumption being that as music is played in the rooms, the absence or boosting of certain tonal elements will detract from the audio quality. The starting point for these previous methods to determine room dimensions, is usually the equation defining the eigenfrequencies within a rigid rectangular enclosure:

$$f = \frac{c}{2}\sqrt{\left(\frac{n_x}{L_x}\right)^2 + \left(\frac{n_y}{L_y}\right)^2 + \left(\frac{n_z}{L_z}\right)^2} \qquad (1)$$

Where n_x, n_y and n_z are integers and L_x, L_y and L_z the length, width and height of the room. Often the best dimensions are given in terms of the ratios to the smallest room dimension. Previous methods for determining room ratios differ, however, in how they utilize Equation (1).

Bolt [1] produced design charts that enabled him to determine good room ratios. His method investigated the average modal spacing to try and achieve evenly spaced modes; the assumption being that if the modal frequencies are evenly spaced, then there will fewer problems with peaks and dips in the modal response. It is now known, however, that using the average mode spacing is not ideal, and the standard deviation of the mode spacing is a better measure. Ratios of 2:3:5 and 1: $2^{1/3}$:$4^{1/3}$ (1:1.26:1.59) are suggested, but Bolt also notes that there is a broad area over which the average modal spacing criterion is acceptable. (Note, this later ratio is often rounded to the commonly quoted figures of 1:1.25:1.6).

Gilford [2] discusses a looser methodology whereby the modal frequencies are calculated and listed. The designer then looks for groupings and absences assuming a modal bandwidth of about 20Hz. The dimensions are adjusted and a recalculation is carried out until a satisfactorily even distribution is achieved. This is a cumbersome process to undertake by hand, but this type of iterative search is easily accomplished using modern computers using numerical optimisation techniques. It is this type of computer controlled optimisation that is set out below as a method for choosing room dimensions. Furthermore, in addition to the use of numerical optimisation to ease the burden of searching, a better basis than modal spacing for evaluating the effects of modes will be detailed. Gilford also states that the 2:3:5 ratio suggested by Bolt is no longer popular and that the axial modes cause the major difficulty in rooms. These points will be returned to later.

Louden [3] calculated the modal distribution for a large number of room ratios and published a list of preferred dimensions based on a single figure of merit. The figure of merit used to judge room ratios is the standard deviation of the intermode spacing, so again this is a regime to achieve evenly spaced modes. The method produces the well known room ratio of 1:1.4:1.9. Louden undertook the investigation by examining 125 combinations of room ratios at a spacing of 0.1. This type of discretized search can limit the potential solutions found. With the optimised techniques developed since Louden published his work, such as the one used below, the search for the best ratios can be undertaken in a more intelligent manner without the need to artificially discretize the ratios tested.

Bonello [4] developed a criterion based on the fact that the modal density should never decrease when going from one third octave band to the next highest band in frequency. Modes with coincidental frequencies are only tolerated in one third octave bands with five or more modes present. Bonello compares his criterion against others used by Knudsen, Olson and Bolt. Justification for his methodology is drawn from his experience as a consultant in 35 rooms.

Walker [5] develops a low-frequency figure of merit based on the modal frequency spacing. The method leads to a range of practical, near-optimum room shapes. Walker discusses how blind application of optimum room ratios does not necessarily lead to the best room, because room quality is volume dependent. The new method outlined in this paper does not use generalised room ratios, and so avoids this problem.

All the above methods have limitations. Equation (1) is only applicable for rigid surfaces. Absorption has a number of effects, for instance it shifts the eigenfrequencies. This is critical for evaluation criteria, as is the case of all the above methods, which examine the modal frequencies or spacing of modes. The new method set out below uses a theoretical model, which although not perfect, is a more accurate model of low frequency room behaviour than Equation (1). Another effect of absorption is that it acts differently on axial, tangential and oblique modes – for example, axial modes will have the greatest magnitude and least damping. None of the above methods account for this fully unlike the new method given below, although Gilford, for example, does discuss the prominence of axial modes. A further difficulty with the above methods is the choice of criterion used for evaluation. For example, Bonello's method makes several assumptions – such as the use of a one-third octave bandwidth, and that five modes in a bandwidth mask the effects of coincident modes – which are empirical rather than fundamental in nature. The new method outlined below acts directly on the modal response of the room, so a criterion based on mode spacing is no longer required. Although an evaluation criterion is still required, as this can be based on the modal response of the room, it is much easier to relate to human perception. This is because the mode spacing is one level more removed from the actual signals received by the listener than the modal response.

Standards and recommendations also stipulate good room ratios for activities such as listening tests and broadcasting. European Broadcasting Union recommendations [6] are discussed by Walker [7]. Walker states that the aim of the regulations appears to be to avoid the worse cases, rather than to provide proscriptive optimum ratios. Consequently, the recommendations cover a wide range of room proportions.

$$\frac{1.1L_y}{L_z} \le \frac{L_x}{L_z} \le \frac{4.5L_y}{L_z} - 4 \quad (2)$$

$$L_x < 3L_z \quad (3)$$

$$L_y < 3L_z \quad (4)$$

In addition, it is stipulated that ratios of L_x, L_y and L_z which are within ±5% of integer values should also be avoid.

The British Standards Institute and International Electrotechnical Commission [8] give slightly different criteria for Equation (2):

$$\frac{L_y}{L_z} \le \frac{L_x}{L_z} \le \left(4, \frac{5L_y}{L_z} - 4\right) \quad (5)$$

The criteria given by Equation (3) and (4) are also stipulated along with recommended floor areas. A recommended room size of 7 x 5.3 x 2.7m (2.59:1.96:1) is given. Older versions of the standard [9] give different recommendations, with a standard room of 6.7 x 4.2 x 2.8m (1.59:1.5:1). These values are also reported in a popular textbook [10].

THE NEW METHOD

The new method is based on producing the flattest possible modal frequency response for the room. It uses an optimising computer algorithm to search for best solutions. First, the prediction models used will be presented, and then the optimising procedure will be discussed.

Prediction Models

For the purposes of this paper, the modal response of the room is defined as the frequency spectrum received by an omni-directional microphone in a corner of the room, when the room is excited by a point source with a flat power spectrum placed in the opposite corner. Two possible models to predict the modal response are considered, a frequency based modal decomposition model and a time based image source model.

Modal decomposition model

The modal decomposition model used is applicable when boundary impedances are large and real. The pressure at **r**(x,y,z) due to a source at $\mathbf{r_o}$(x_o,y_o,z_o), at an angular frequency ω, is given by [11]:

$$p(\bar{r},\omega) = \sum_{n_x}^{\infty}\sum_{n_y}^{\infty}\sum_{n_z}^{\infty} \frac{A_{\omega n}(\bar{r},\bar{r}_0)}{(\omega^2 - \omega_n^2 - j2\omega_n\delta_n\omega)} \quad (6)$$

where:

$$A_{\omega n}(\bar{r},\bar{r}_0) = j\rho\omega S_0 c^2 p_n(\bar{r})\, p_n(\bar{r}_0) \quad (7)$$

$$p_n(\bar{r}) = \cos(k_{nx}x)\cos(k_{ny}y)\cos(k_{nz}z) \quad (9)$$

$$p_n(\bar{r}_0) = \cos(k_{nx}x_0)\cos(k_{ny}y_0)\cos(k_{nz}z_0) \quad (10)$$

$$\omega_n / c = k_n = \sqrt{k_{nx}^2 + k_{ny}^2 + k_{nz}^2} \quad (11)$$

$$\delta_n = \frac{c}{\omega_n}\left(\frac{\varepsilon_{n_x}\overline{\beta}_x}{L_x} + \frac{\varepsilon_{n_y}\overline{\beta}_y}{L_y} + \frac{\varepsilon_{n_y}\overline{\beta}_y}{L_y} \right) \quad (12)$$

where $\overline{\beta}_x, \overline{\beta}_y$ and $\overline{\beta}_z$ are the average admittance of the walls in the x, y and z directions

$$\begin{aligned} \varepsilon_{n_x} &= 1 \; for \; n_x = 0 \\ &= 2 \; for \; n_x > 0 \end{aligned} \quad (13)$$

$$k_{nx} = \frac{n_x \pi}{L_x} \quad (14)$$

Similar expressions for ε_{ny}, ε_{nz}, k_{nx} and k_{ny} are used. ρ is the density of air, S the surface area of the room, V the volume, c the speed of sound.

Image source model

The image source model is a fast prediction model for a cuboid room. The image solution of a rectangular enclosure rapidly approaches an exact solution of the wave equation as the walls of the room become rigid. The energy impulse response is given by:

$$E(t, \overline{r}, \overline{r}_0) = \sum_{n_x}^{\infty}\sum_{n_y}^{\infty}\sum_{n_z}^{\infty}\sum_{i=1}^{2} R_{n,i}^2 \frac{1}{d_{n,i}^2} \quad (15)$$

$$d_{n,i} = \sqrt{d_{n_x,i}^{\;2} + d_{n_y,i}^{\;2} + d_{n_z,i}^{\;2}} \quad (16)$$

$$\begin{aligned} d_{n_x,1} &= (L_x - x_0) + (-1)^{n_x} x \\ &\quad + L_x(n_x - 1) + L_x((-1)^{n_x+1} + 1)/2 \end{aligned} \quad (17)$$

$$\begin{aligned} d_{n_x,2} &= x_0 + (-1)^{n_x+1} x \\ &\quad + L_x(n_x - 1) + L_x((-1)^{n_x} + 1)/2 \end{aligned} \quad (18)$$

Similar expressions for the distances in the y and z direction are used. The surface reflection factors are given as follows:

$$R_{n,i} = R_{x,i} R_{y,i} R_{y,i} \quad (19)$$

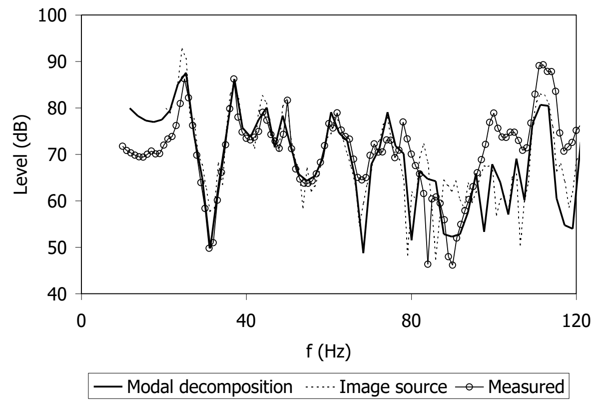
$$R_{x,i} = R_{x,i}^{|n_x|} R_{x,\mathrm{mod}(i,2)+1}^{|n_x - 1|} \quad (20)$$

Where $R_{x,1}$ and $R_{x,2}$ are the surface reflection factors for the front and rear walls respectively, and similar expressions for the distances in the y and z direction are used. Reflection factors are approximated to be purely real. Once the energy impulse response is obtained, this is Fourier Transformed to form the modal frequency response. For soft walls, the image source construct becomes less accurate as representing the image sources as pure point sources is no long applicable. These inaccuracies become greater as the reflection orders increases.

Prediction model critique

Both the modal decomposition and image source models offer a better representation of the sound field in the space than the simple modal frequency equation shown in Equation (1). This is primarily because the modal decomposition and image source models allow for absorption, but also because it is possible to calculate a quantity - the modal response - that is easier to relate to the listener experience. Both models, however, are not completely accurate. Figure 1 compares measurements in a listening room to the modal decomposition and image source models. The listening room has dimensions 6.9 x 4.6 x 2.8m. All the walls were smooth plastered concrete except the back wall which was covered with diffusers, some diffusers were on the ceiling and the floor which was covered with carpet.

Figure 1. Prediction models compared to a measurement in a listening room.



Below 100 Hz, good agreement between the models and the measurement are shown. The agreement diverges above 100Hz – see below. Slightly better agreement can be achieved [12] by taking more terms in Equations (6) and (15). The models deliberately used a reduced number of terms in the infinite sums to enable calculations to be quick enough for subsequent optimisation.

Great care to normalise for loudspeaker resonance is required for these measurements, in this case the loudspeaker resonance was about 80Hz. The sound power of the loudspeaker is difficult to measure, as anechoic conditions are not achieved at 20Hz in normal test chambers. The cone acceleration was used as a reference for the frequency response normalisation. This was measured by an accelerometer attached near the centre of the loudspeaker cone. If the cone radiates as a piston at such low frequencies, the free-field pressure should be omni-directional and proportional to the cone acceleration. Thus, the cone acceleration provides a convenient means of normalisation. This worked well over most of the frequencies from 30 to 100Hz. At the lower and upper ends the normalisation might still be affected by poor signal to noise ratio and directional radiation form the cone. Indeed, it is assumed that the divergence between the measurement and predictions above 100Hz is due to cone directivity.

Consequently, the method for choosing room dimensions is based on a better prediction model than previous methods. There is, however, scope for future improvement, by including better prediction models when they are developed. There are some basic problems with both the modal decomposition and image source models, and currently there are no established solutions to deal with these difficulties. For example, while absorption coefficients for surfaces are widely available, the surface impedances, which includes both phase and magnitude information, are not. Indeed, given that room surfaces at low frequencies will often not behave as isolated local reacting surfaces, defining a surface impedance can be problematical. Consequently, for this work an assumption of no phase change on reflection has to be made, which means that the models are more accurate for walls that are more rigid. It might be envisaged that a finite element model could overcome some of these difficulties, but currently the calculation time would be too slow for optimisation. During an optimisation process, many hundreds or thousands of room configurations have to be calculated, consequently the prediction time for a single calculation must be kept small.

For the results presented here the image source model was favoured over the modal decomposition model. This is because the image source model is considerably faster. For the modal decomposition model, all modes within the frequency range of interest must be considered, plus corrections for those outside the range must be done [13]. In the image source model, all images contribute to the impulse response in a cuboid room. Consequently, using the image source model reduces the optimisation time. (Furthermore, the methodology outlined has been previously applied to finding the best location for loudspeakers and listeners in rooms [14] and using a time based approach for that problem enables the early arriving sound to be examined as well as the modal response). The relationship between the modal decomposition and the image solutions for a loss-less room has been derived and shown to be equivalent for a rigid boundary [15].

Optimisation Procedure

Numerical optimisation techniques are commonly used to find the best designs for a wide variety of engineering problems. In the context of this paper, a computer is used to search for the best room dimensions. The iterative procedure is illustrated in Figure 2. The user inputs the

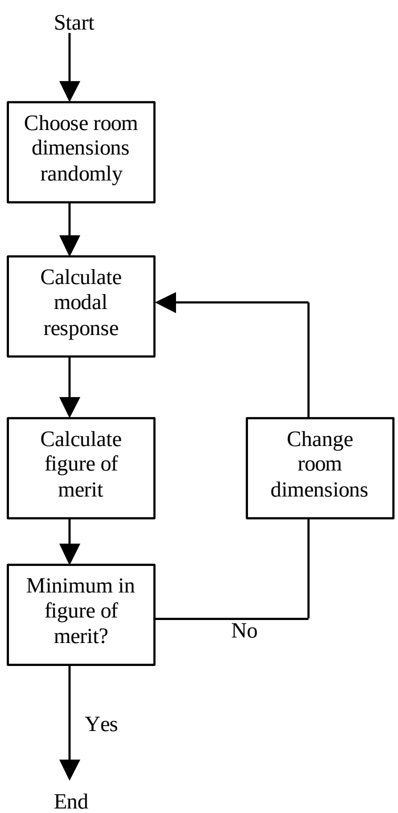


Figure 2. The optimisation procedure

minimum and maximum values for the width, length and height, and the computer finds the best dimensions within these limits. The computer predicts the modal response of the room and rates the quality of the spectra using a single figure of merit (cost parameter). The computer then tries other dimensions trying to find those that have the lowest figure of merit. A completely random search is too time consuming, and so one of the many search algorithms that have been developed for engineering problems was used. In this case, a simplex method was used [16], which is not the fastest procedure but is robust and does not require knowledge of cost parameter derivatives.

In developing a single figure of merit it is necessary to consider what would be the best modal response. It is assumed that the flattest modal response corresponds to the ideal. This is done even though a perfectly flat response can never be achieved, as in the sparse modal region there will always be minima and maxima in the frequency response. The cost parameter used is the squared deviations of the modal response from a least squares straight line drawn through the spectra. If the modal response level of the n^{th} frequency is $L_{p,n}$ then the cost parameter ε is:

$$\varepsilon = \sum_{n=1}^{N} (L_{p,n} - mf_n + c)^2 \qquad (21)$$

Where m and c are the gradient and intercept of the best fit line and the sum is carried out over n frequencies, f_n. This is illustrated in Figure 3. Consequently, this is a least squares minimization criterion,

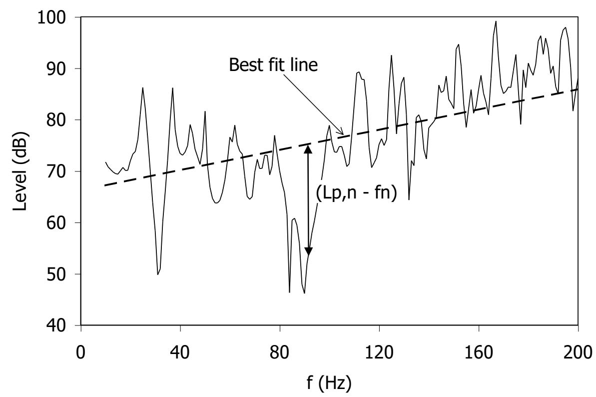


Figure 3. Use of best fit line to get figure of merit (no spectrum smoothing).

which is commonly used in engineering. The deviation from a best fit line rather than the mean is used because it is assumed that slow variation in the spectrum can be removed by simple equalisation, and what is important is to reduce large local variation. Before calculating Equation (21), some smoothing over a few adjacent frequency bins is used. This is done to reduce the risk of the optimisation routine finding a solution that is overly sensitive to the exact room dimensions. Furthermore, in prediction models very exact minima can be found which would never be replicated in real measurements; the smoothing helps mitigate against this.

Test bed

The optimiser was run for a wide variety of room sizes: 7m ≤ L_x ≤ 11m, 4m ≤ L_y ≤ 8m and 3m ≤ L_z ≤ 5m. Two hundred solutions were gathered. In most multi-dimensional optimisation, repeated running of the procedure from random starting positions will give different "optimum" solutions. This happens because the optimising algorithm gets stuck in a local minimum that is not the best solution available - the so-called global minimum. A large number of solutions were gathered to enable the performance of the optimisation to be tested and a statistical analysis of the solutions found to be undertaken. If used as a design tool, far fewer solutions could be found, and the best used. The best solutions found were compared to the previously known best ratios outlined above.

A frequency range of 20-200Hz was chosen, as the flatness of the modal response was not particularly sensitive to dimension changes above 200Hz. As might be expected, the gains to be made in avoiding degenerate modes are at lower frequencies where the modes are relatively sparse. The frequency range for optimisation may also be guided by the Schroeder frequency.

Results

To compare to previous work, the best solution whose volume was roughly the same to that used by Louden was chosen. This is to enable a fair comparison. Absorption coefficients were chosen to be typical of those found at low frequency in listening rooms.

Figure 4 shows the new optimised modal response compared to one of the ratios suggested by Bolt, 2:3:5. In addition, the modal spectrum for the worst dimensions found during the search is shown to give a sense of the range of spectra that can be achieved. (The worst case had a ratio 1:1.075:1.868). As expected, a completely flat spectrum is not achieved with optimisation. Clear improvement on the Bolt 2:3:5 room is seen, however. The 2:3:5 rooms suffers from significant dips, for example at 110Hz.

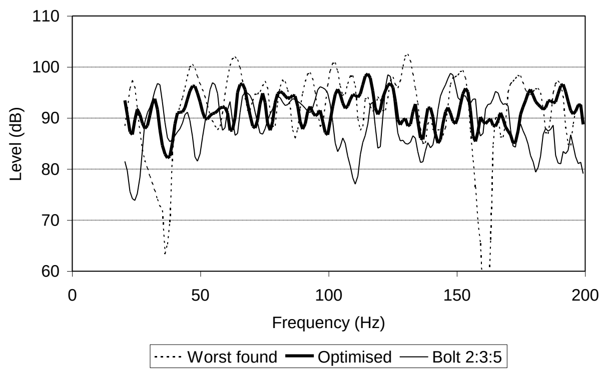


Figure 4. Modal response for three room dimensions including Bolt's 2:3:5 ratio.

The best ratio found by Louden, 1:1.4:1.9, is compared to the new optimised response in Figure 5. Improvement on the Louden ratio is achieved, although the improvement is less marked than with 2:3:5. Bolt also suggested the ratios 1:1.25:1.6, which also meets Bonello's criteria. Figure 6 shows the spectra compared to the optimised solution. The modal response spectrum achieved by optimisation is clearly flatter.

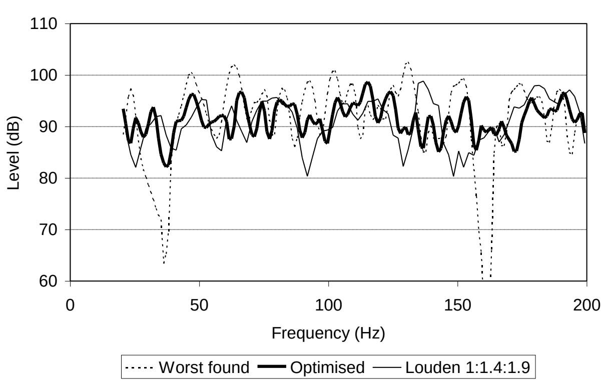


Figure 5. Modal response for three rooms, including the best ratio found by Louden 1:1.4:1.9.

The optimised solution was also compared to the regulations and standards mentioned above. All of the ratios by Bolt, Bonello and

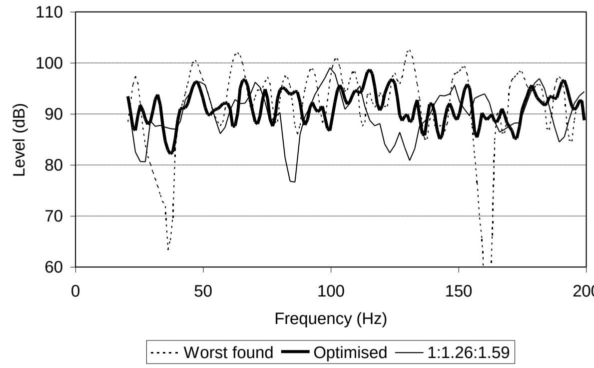


Figure 6. Three modal spectra including 1:1.26:1.59.

Louden presented before pass the EBU and IEC regulations as does the best optimised solution. Only the worst case fails to meet the regulations. The standards appear to achieve their remit of not being overly proscriptive while avoiding the worst cases.

A comparison with the preferred standard room sizes given in the standards and regulations was also undertaken. Figure 7 compares the optimised solution to the old and new IEC regulations. The new standard room and the optimised solution are very similar in performance. While the cost parameter for the optimised solution is better (2.2) than for the new standard room (2.5), this does not translate into an obvious improvement in the spectra. (This gives a little evidence for the sensitivity of the cost parameter; the difference limen appears to be greater than 0.3). The old standard room, however, is far from optimum, indicating a wise revision of the standard.

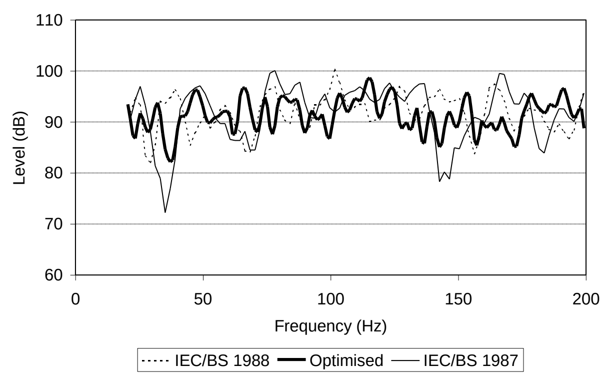


Figure 7. Comparison of optimised solution with standard rooms.

Finally, the optimised solution is compared to the "golden ratio" in Figure 8. The golden ratio is often quoted in the audio press, and so is of interest. It was tested for this reason, despite the fact that the rationale behind the golden ratio for room dimensions does not appear particularly compelling from a scientific point of view. It can be seen that the optimised solution has a more even modal response and so is better.

Discussions

The new method produces as good or better room dimensions than those based on previous work. The new method has been shown to be an efficient way of finding optimum dimensions. The modal spectrum in a room is complex, and there does not appear to be one set of magical dimensions that significantly surpass all others in performance. There may be a numerically global minimum, but many of the local minima in the optimisation problem are actually equivalent in terms of the modal response achieved. One significant advantage of this optimisation technique is that it is possible to incorporate constraints that may happen in real buildings. For example, if the height of the ceiling is fixed in the building, then it can be fixed in the optimiser, which can then look for the best room width and length within constraints given by the user.

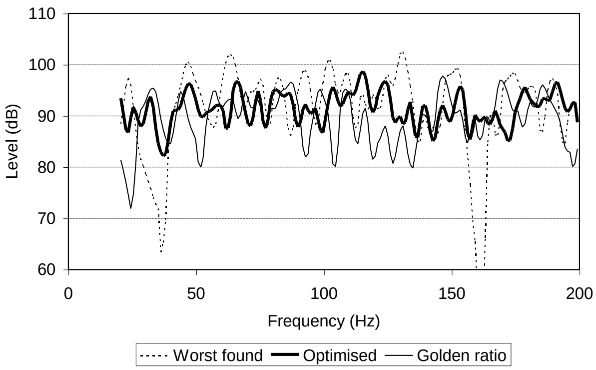


Figure 8. Three room modal responses compared including the "golden ratio"

Conclusions

A method has been presented to enable the size of small critical listening spaces to be determined. The criteria for room size being to minimise the coloration effects of low frequency modes monitored by a predicted modal response. The method is an improvement on previous methods in that the theoretical basis is a more accurate model of the room response than examining the modal spacing in frequency for a rigid box. The system is flexible in that it can search for the best dimensions within constraints set by the designer. Furthermore, the procedure has flexibility in that as better prediction models of rooms become available, they can be used in the general optimisation design procedure. Results demonstrate that the new search method produce room sizes that match or improve on the room ratios published in literature.

Acknowledgements

The authors would like to thank Professor Lam for his advice on prediction models, and Mark Avis for allowing us to use his measurements.

References

1. R.H.Bolt. Note on the normal frequency statistics in rectangular rooms. J.Acoust.Soc.Am. 18(1) 130-133. (1946)
2. C.L.S.Gilford. The acoustic design of talk studios and listening rooms. J.Audio.Eng.Soc. 27. 17-31. (1979).
3. M M Louden. Dimension ratios of rectangular rooms with good distribution of eigentones. Acustica. 24. 101-104 (1971).
4. O.J.Bonello. A new criterion for the distribution of normal room modes. J.Audio.Eng.Soc. 29(9) 597-605 (1981). Note erratum given in 29(12) 905 (1981).
5. R. Walker. Optimum Dimension Ratios for Small Rooms. Preprint 4191. 100th Convention of the AES. (5/1996).
6. EBU Technical Recommendations R22-1998. Listening conditions for the assessment of sound programme material.
7. R.Walker. A Controlled-reflection listening room for multichannel sound. proc. IoA(UK). 20(5) 25-36. (1998)
8. IEC 60268-13, BS6840 13, Sound system equipment – part 13 listening tests on loudspeakers. (1988).
9. IEC 268-13:1985, BS6840 13 (1987), Sound system equipment – part 13 listening tests on loudspeakers.
10. Loudspeaker and Headphone Handbook" 2nd Edition. Edited by John Borwick. Focus Press, ISBN 0-240-51371-1.
11. P M Morse and K U Ingard, "Theoretical Acoustics" McGraw-Hill, 555-572, (1968).
12. Y.W. Lam. An Overview Of Modelling Techniques For Small And Large Performance Spaces. Proc. Institute of Acoustics (UK). 22(2), Pages: 297-304 (2000).
13. J.B.Allen and D.A.Berkley, "Image method for efficiently simulating small room acoustics" J.Acoust.Soc.Am. 65(4), 943-950 (1979).
14. T. J. Cox and P. D'Antonio. Room Optimizer: A Computer Program to Optimize the Placement of Listener, Loudspeakers, Acoustical Surface Treatment, and Room Dimensions in Critical Listening Rooms", proc. 103rd Convention of the Audio.Eng.Soc, Preprint 4555, Paper H-6, New York (1997).
15. C L Peckeris. "Ray theory versus normal mode theory in wave propagation problems," proceedings of the symposium in applied mathematics, II, 71-75 (1950).
16. W. H. Press et al. "Numerical Recipes, the Art of Scientific Computing" Cambridge University Press. 289-292. (1989).