Audio Engineering Society

Convention Paper

Presented at the 126th Convention
2009 May 7–10 Munich, Germany

Subjective Validity of Figures of Merit for Room Aspect Ratio Design

Matthew Wankling[1], Bruno Fazenda[1]

[1] *University of Huddersfield, Huddersfield, HD1 3DH, United Kingdom*

Correspondence should be addressed to M. Wankling (`m.wankling@hud.ac.uk`)

ABSTRACT

Attempts have long been made to classify a room's low frequency audio reproduction capability with regards to its aspect ratio. Common metrics used have relied on the homogeneous distribution of modal frequencies and from these a number of 'optimal' aspect ratios have emerged. However, most of these metrics ignore the source and receiver coupling to the mode shapes - only a few account for this in the derivation of a figure of merit. The subjective validity of these attempts is tested and discussed. Examples are given of supposedly good room ratios with bad performance and vice versa. Subjective assessment of various room scenarios is undertaken and a ranking order has been obtained to correlate with a proposed figure of merit.

1. INTRODUCTION

It has long been desirable to reproduce audio at the highest possible quality in a variety of locations. Recording studios, home living spaces and even automobile cabins are all places where the audio industry has focused on producing a better quality of reproduced sound.

In these locations, it is not only the quality of the audio components which will influence the perceived quality of sound. The physical surroundings also play an important role. Of particular importance is the reproduction of low frequencies. In small listening spaces, where low frequency wavelengths are comparable in size to the dimensions of the room, standing waves, or room modes, are set up. Depending on the room dimensions and conditions, these modes are created at differing frequencies, and have a number of related parameters. The addition of many modes creates a unique frequency response, not only for that room but for each listening position within it. This unique response often leads to degradation of audio playback. Therefore, there has been a significant amount of research in this area, in an attempt to optimize the physical sound-field so as to have the least affect upon the listener. A number of optimization techniques have been attempted, such as low frequency absorption, careful positioning of the loudspeakers in the room [1], a number of equalization techniques [2] and also the addition of multiple subwoofers throughout the room [3]. However, one of the earliest, and perhaps on the face of it, simplest methods of optimization, was that of designing the listening room dimensions according to a carefully considered aspect ratio. Although so called golden ratios have been questioned, this is still a theory that is widely considered. One consistent problem with room acoustic optimization studies is the reliance on theoretical calculations. Whilst certainly a good place to start, if objective measures calculated cannot be shown to correlate with subjective testing, the validity of using them to produce practical design recommendations must be questioned. This paper discusses the theoretical nature of optimization by room aspect ratio and presents a psychoacoustic listening test to determine if the scoring systems applied to rooms based on these ratios have a subjective relevance.

2. CLASSIFICATION METRICS

A classification metric is a method of quantifying the quality of low frequency reproduction in a room according to some objective measure. This metric should of course, relate to the subjective perception of low frequency sound. Historically, metrics have been produced which produce differing scores dependant on the ratio of the three room dimensions.

Let us briefly consider why they came about, and how they are calculated. Here we turn to two related areas modal spacing and frequency response.

2.1. Modal Spacing

The idea of optimizing room dimensions comes about due to the modal frequencies relationship to each dimension. As previously discussed, modes are set up as standing waves, and occur at specific frequencies depending on the dimension (equation 1).

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

Where Lx Ly and Lz are the three room dimensions, c the speed of sound and n an integer depending on the number of the mode.

The combination of many modes results in an overall frequency response. Audible problems are generally thought to occur where large peaks or dips occur in this response. When these correspond to the frequencies present in the audio, degradation is heard, often in the form of uneven bass and long 'ringing' decays. Therefore, it has long been considered that a flat frequency response offers the ideal listening conditions. A room response effectively flattens out as the Q factor of the modes producing that response tends towards zero. This is directly associated with the damping of energy in the modes. A lower Q has been shown to reduce the audible effects of modes [4,5]. Here it was observed that the spacing between modal frequencies may become important. If two modes occur at the same frequency, a case known as modal degeneracy, the summation will cause a stronger peak at this frequency. Conversely, if these two frequencies are spread apart, the overall response of the two will tend towards flat (see Figure 1). Two modes may share the same frequency if two room dimensions are the same. The basis of aspect ratio optimization lies in the reduction of this scenario.

Bolt was one of the first to optimize in this way, looking at those aspect rations which would cause least modal degeneration, finally determining a 'blob' on a map within which acceptable ratios were supposed to lie [6]. Others followed, with Louden publishing a list of best ratios according to a metric based on the standard deviation of modes from a statistically generated ideal [7]. The Bonello Criteria, named after its author in the 1980's, stated a number of modal spacing conditions which should be met in order to create the best responses [8]. Walker [9] and Gilford [10] also suggested methods based upon the spacing of modal frequencies. Whilst authors have always taken care to state that their ratios or criteria's should not be followed blindly, they continue to permeate in the minds of many.

2.2. Frequency Response Deviation

Each of the above methods contains an underlying problem. Whilst they may seem to deal with the problem at hand, and flatten the room's frequency

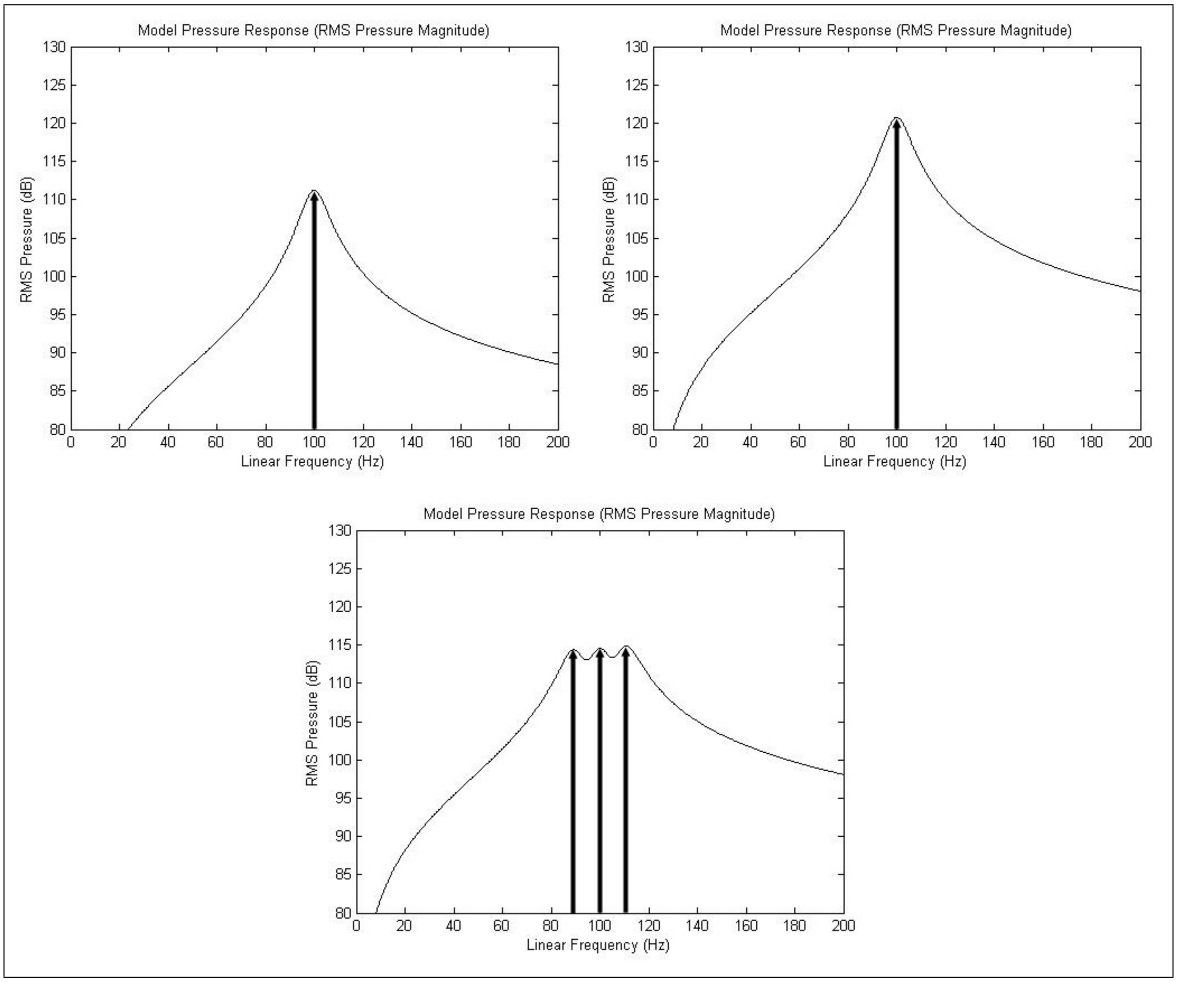

Fig. 1: The effect of multiple modes on frequency response. a) single mode at 100Hz b) three modes each at 100Hz c) three modes at 90Hz, 100Hz, 110Hz

response to give a reasonable idea of the room's performance, it must be noted that this theory does not account for the modal interaction within the room. In reality, the pressure of each mode will vary through positive and negative values depending on position within the enclosure. This is usually referred to as 'mode shapes'. This means that the frequency response cannot be assumed in the simplistic terms shown in Figure 1, as position of both source and receiver may account for differing phases at a single point in the enclosure and for a particular modal interaction dependent on that. This can severely alter the response. In each of the metrics above based on modal spacing, no room position is accounted for. In response to this problem, the work of Cox et al. created a frequency response using a computer model [11]. With the actual response available, a figure of merit can then be obtained by studying how it deviates from a desired response, say that of a flat response.

Again assumptions underpin this type of figure of merit (FOM). It must be assumed that subjectively, we prefer the resulting audio if on average it lies closer to a flat magnitude frequency response. This may not necessarily be the case. Previous work suggests that the evidence of peaks and dips at specific frequencies should not be ignored, playing a much greater part in our perception of audio quality than averaging techniques account for [5]. To highlight this problem, consider an artificial case where a single resonance exists, at 100Hz. When this is modeled, and a FOM obtained using the deviation from a smooth response, a relatively high score of 0.7 is achieved (Figure 2). However, listening to a sound source in this room is clearly undesirable, as the presence of such a resonance is immediately detectable.

In order to study the subjective relevance of these FOMs further, a psychoacoustic listening test was conducted which identifies subjects preference to a number of auralisations. Virtual rooms were modeled, allowing cases to be chosen according to a FOM similar in nature to that of Cox et al. [11]

It can be shown that, when deriving a FOM using the deviation of a modeled frequency response from a smooth curve, good scores can be obtained in supposedly poor rooms according to modal spacing metrics. These two types of metrics have previously been contrasted by Fazenda et al. [12]. This test will therefore also attempt to comment on the relevance of using such spacing metrics.

3. EXPERIMENTAL SETUP

3.1. Auralisation

In order to test a variety of spaces, virtual room auralisations have been produced, allowing the comparison of multiple cases quickly over high quality headphones. The authors have previously used this technique with success [4,5,12].

An audio sample is used which contains sufficient low frequency energy, and this is first convolved with a real recorded binaural impulse response. This places the sample within the real room. This signal is then high pass filtered to remove frequencies below 250Hz. This high frequency region is kept constant across all test cases. The low frequency region can then be replaced by a modeled room, generated using the modal decomposition method, shown in Eq. 2. The model generates the modal response up to 300Hz. Whilst mostly interested in the modes up to 250Hz, this additional 50Hz allows the lower frequency residues from modes in the region 250-300Hz to influence the response below 250Hz.

$$P_w(r) = jw\rho Qc^2 \sum_n \frac{p_n(r)p_n(r_0)}{(w^2 - w_n^2 - 2jd_nw_n)X_n} (Eq.2)$$

Finally, a low pass filter is placed across this low region. The high and low pass filters combine in an eighth order Linkwitz-Riley configuration, producing steep roll off characteristics whilst ensuring that the resulting sample contains no irregularities at the crossover frequency.

3.2. Low Frequency FOM

As previously mentioned, the low frequencies were modeled using the modal decomposition method. The decay time of all mode types (axial, tangential and oblique) was assumed to be equal, modeled with an exponentially decaying reverb time relating to that in the real room used in the auralisation of high frequencies. A sampling frequency of 5000Hz was used to ensure that the resolution was high enough and no ringing or aliasing was present in the output signal. The Figure of Merit (FOM) can then be directly produced from the first 300Hz of this low frequency region. The decomposition model outputs a complex frequency response, from which the RMS magnitude is taken. A third order polynomial line of regression is fitted to this, producing a smooth curve through the response.

This smooth response therefore closely resembles the situation where the decay time is theoretically zero, whilst still retaining the overall magnitude deviances in frequency for that particular room. The difference between the RMS magnitude response and smooth curve (see Figure 3) is then calculatd at each sampling point, and the mean taken to produce the FOM itself. A perfect replication of the curve results in a score of 1, and the worst case scenario (FOM = 0) would occur if the average difference was 10dB.

3.3. Test Cases

To investigate the reliability of this FOM in determining rooms with subjectively good audio reproduction, a number of room auralisations were produced, each scoring differently according to the FOM under study. Three different room scenarios were used. This allows us to test the FOM within:

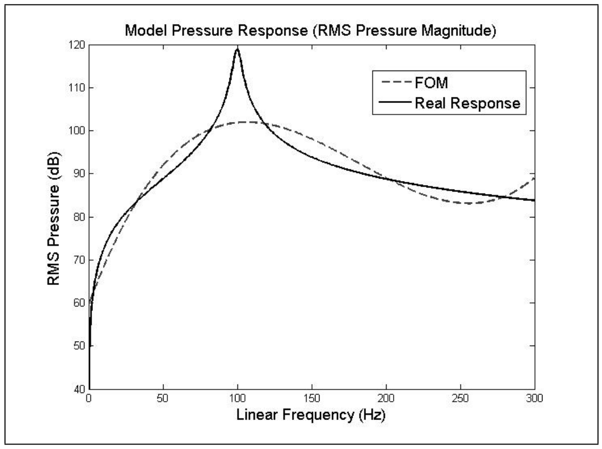

Fig. 2: A single resonance at 100Hz, with a smooth curve fitted through. The FOM score in this case is relatively high - 0.70 (FOM from 0-1)

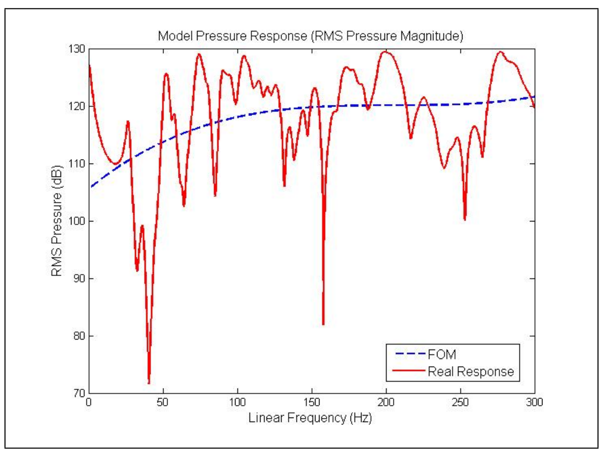

Fig. 3: Frequency Response and Figure of Merit Regression line for Case 5

A. A theoretically bad room - cuboid

B. A room which should sound good according to frequency spacing metrics

C. The same good room excited at a different source position

Finally, a reference sample is produced from the smooth fitted line. In theory, this reference becomes the benchmark by which the other FOM scores are rated, and should theoretically be the best sounding room, as the decay time is practically zero and there are no obvious peaks or dips in the response. When producing a map of FOM's across a sample of positions within the room, it becomes apparent that a receiver position right next to the source scores highest. This is to be expected as the direct sound dominates in this position. Although unrealistic, the FOM score is highest, so a sample of this type is also included for comparison.

In order to determine room positions to test, FOM maps across one place in the z dimension were generated. Figures 8-10 (Appendix 1) show examples of this.

Fixed Room Parameters				Test Case	x Receiver Position	y Receiver Position	FOM
Room A	x=4.6	y=4.6	z=4.6	1	0.4	0.4	0.78
	xSource=0.2	ySource=0.2	zSource=0.2	2	1.0	4.0	0.38
				3	2.8	2.8	0.51
Room B	x=4.69	y=6.36	z=3.35	4	3.0	4.8	0.72
	xSource=0.2	ySource=0.2	zSource=0.2	5	2.4	3.8	0.34
				6	3.6	1.8	0.58
				10	0.4	0.4	0.81
Room C	x=4.69	y=6.36	z=3.35	7	1.0	0.2	0.76
	xSource=0.9	ySource=0.5	zSource=1.3	8	3.4	3.2	0.40
				9	2.2	3.0	0.62
Ref	x=4.69	y=6.36	z=3.35	11	0.001	0.001	0.99
	xSource=0.001	xSource=0.001	zSource=0.001				

Table 1: Room Parameters and FOMs for the 11 Test Cases (zReceiver=zSource)

3.4. Test Interface

The test itself was created in Matlab [13]. A graphical user interface was produced, consisting of play buttons for each of the 11 cases. Each button is accompanied by a slider which is to be placed by the subject. Subjects were instructed simply to "rate the following audio samples with relation to their low frequency reproduction quality". No explanation of the characteristics of this 'good quality reproduction' was given. This naturally leads to a preference test. It was decided that this best allows the determination of subject's preference, rather than testing their ability to identify characteristics which have been pre-defined.

For each test the 11 cases were randomized so that the play order would not be a biasing factor. Subjects were free to repeat the cases and alter their score as often as they liked until happy. In total, 15 subject's responses were recorded, and each subject rated the same 11 cases with three different musical samples. These were short clips of average 6 seconds, chosen specifically to excite low frequency energy whilst not being too complex in nature. Sample 1 is a short refrain with most of the low frequency energy in the drum track, and a few sustained bass notes. Sample 2 is a short double bass riff with a number of notes while sample 3 is a modern remix of a jazz track, which has clean drum and bass notes.

4. RESULTS

Results are shown and statistical analysis has been carried out to reveal their significance.

Of primary concern is the relationship between the 11 test cases and the subjective scores given to them. It can be noted, based on a high standard deviation of data and anecdotal evidence collected from subjects after the test that agreement as to the quality of low frequency reproduction was not always found - judgement of good quality bass reproduction would appear variable. Furthermore, as the test was in essence a preference test most subjects were found to, in practical terms, apply ranking scores, refining their scores comparatively with similar sounding cases.

For this reason, each individual subjects scores have been normalized, thereby maximizing each scale between their minimum and maximum scores. Table 2 shows the 11 cases and mean normalized subjective scores for the three music samples, along with the standard deviation.

Figure 4 presents this information graphically with case arranged in ascending order in terms of calculated FOM. Here, as the subjective scores have been normalised, so too must the FOM. Although differences are apparent, it can be stated that, on average, some trend is visible - increasing FOMs result in an

		Case										
		1	2	3	4	5	6	7	8	9	10	11
	FOM	0.68	0.06	0.26	0.58	0.00	0.37	0.65	0.09	0.43	0.72	1.00
Sample 1	Mean	0.59	0.41	0.20	0.53	0.53	0.43	0.45	0.38	0.41	0.69	0.78
	St.Dev.	0.24	0.31	0.28	0.29	0.31	0.36	0.32	0.28	0.26	0.33	0.27
Sample 2	Mean	0.59	0.22	0.34	0.49	0.26	0.45	0.51	0.25	0.57	0.51	0.53
	St.Dev.	0.21	0.32	0.35	0.31	0.25	0.26	0.43	0.30	0.27	0.35	0.391
Sample 3	Mean	0.45	0.13	0.18	0.39	0.25	0.32	0.68	0.32	0.21	0.66	0.72
	St.Dev.	0.31	0.18	0.18	0.33	0.31	0.24	0.39	0.32	0.21	0.26	0.34

Table 2: Normalised mean and standard deviation for each of the three music samples. FOM also normalised

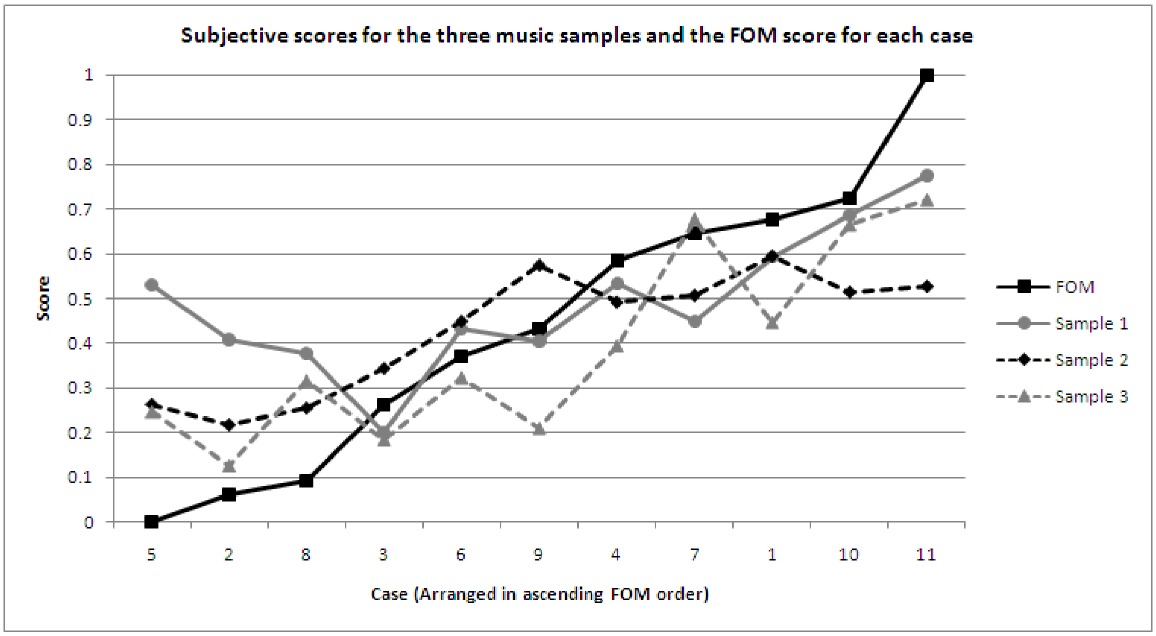

Fig. 4: Normalised mean subjective scores for the three music samples compared with the FOM. Case arranged in order of ascending FOM

increased subjective score.

Statistical analysis was also carried out upon the normalized data set. Firstly, a two way ANOVA was run. This takes the two independent variables, 'case' and 'music sample' and returns a value which shows acceptance or rejection of a null hypothesis. Here the null hypothesis is that neither music sample nor case affect the subjective rating. The ANOVA result is shown in Table 3. For this analysis, the two unrealistic references cases are omitted, as these are highly likely to score significantly different and could therefore confound ANOVA interpretation.

The probabilities of 0 and 0.0001 for the columns (musical sample) and rows (test case) both reject the null hypothesis, suggesting both factors as highly significant. It is of course expected that the test case will be so, in accordance with the change in room frequency response with each different case. However, that musical sample is considered highly significant is interesting. A visual inspection of Figure 4 would

Source	SS	df	MS	F	prob>F
Columns	6.5095	8	0.81369	8.08	0
Rows	1.8451	2	0.92254	9.17	0.0001
Interaction	3.1564	16	0.19728	1.96	0.0148
Error	38.046	378	0.10065		
Total	49.557	4.4			

Table 3: ANOVA Results

suggest that the subjective scores lie within a similar region, especially considering the high standard deviation. Furthermore, previous studies have suggested that musical style is an insignificant factor in modal listening tests [12].

An explanation for this may be found when looking at the nature of ANOVA. With this procedure, the null hypothesis is rejected if a single element shows a significant variation from the others. Therefore, one musical sample may define this result if it differs from the others. To further test the significance of individual cases and music sample, a series of t-tests were carried out. The two sample t-test looks at two sets of data (here, subjective scores according to two different musical samples) and determines whether or not the scores are significantly different. Comparisons were made as follows: In each of the 11 test cases, between scores given for music sample:

i. 1 and 2

ii. 1 and 3

iii. 2 and 3

The return variable from the t-test function in Matlab is a probability that the scores come from the same population. If this is below 5%, Matlab returns a 1, showing the music sample was significant in judging the cases quality. Conversely, if above 5%, a zero is returned and the music sample is not considered significant. Table 4 summarises the results.

As can be seen, the musical sample is significant in four of the 33 cases. Interestingly, where it is shown to be so, the case number is similar. Cases 2, 5 and 9 all appear to show that the music may play a significant role in the score given. This result is discussed further in section 5.

With these t-tests suggesting that in most cases, the musical sample does not affect the score, we can take a mean across the three samples. This is considered useful as, typically, listening spaces such as the ones modelled are used for playback of a wide variety of subject material. For this reason, it is interesting to note the overall mean score across three rooms. Figure 5 shows this, again using normalised data. A general trend is visible, with differences becoming apparent at the extremities.

As previously mentioned, in a practical sense, the test was often one of ranking the cases, with subjects placing the sliders on the GUI in relation to whether the case sounds better or worse than similar ones. It is therefore interesting to take the scores given by each subject, and rank them 1 to 11. Results are plotted in Figure 6. Here, four tiers are categorised. A 'Bad' score is registered when a subject ranks the case as one of the three worst; 'Medium' when the case in ranked in positions four, five or six; 'Good' when ranked in position seven, eight or nine, and finally 'Best' when ranked as one of the top two cases (positions 10 or 11). The number of subjects who ranked the cases in each particular category is denoted by the size of the circle. The cases on the horizontal axis are again arranged in ascending FOM order.

Again a general trend is visible, and this suggests that the FOM is reasonably accurate in predicting the preference of certain cases over others. This is certainly true of the higher scoring FOMs. However, in a similar fashion to when comparing the actual means, there is some disagreement in the lower Figure of Merit cases.

Finally, it is interesting to compare the results between the samples in the three test rooms. It becomes apparent that the subjects have not rated

Comparison	1	2	3	4	5	6	7	8	9	10	11
1&2	0	0	0	0	1	0	0	0	0	0	0
1&3	0	1	0	0	1	0	0	0	0	0	0
2&3	0	0	0	0	0	0	0	0	1	0	0

Table 4: t-test Results

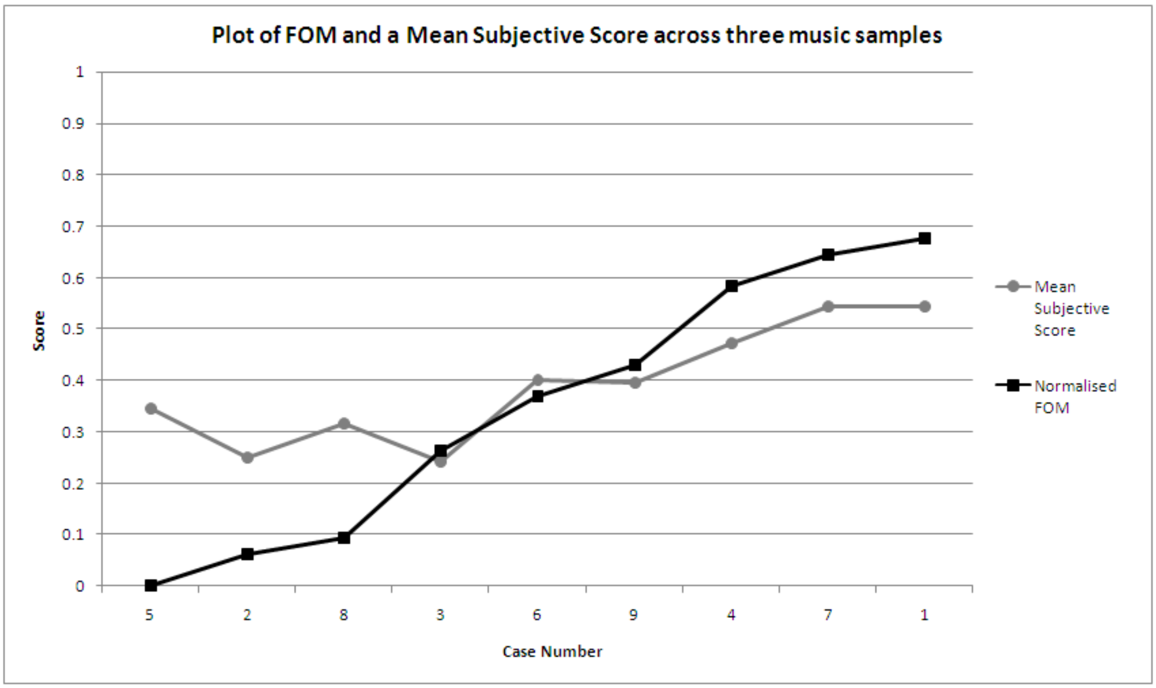

Fig. 5: Normalised FOM compared to subjective scores, averaged across the three music samples (Reference samples 10 and 11 are omitted)

	Room A			Room B			Room C		
FOM	0.38	0.51	0.78	0.34	0.58	0.72	0.40	0.62	0.76
Difference between FOM and Subjective Mean	-0.08	-0.22	-0.23	+0.03	-0.16	-0.26	-0.05	-0.22	-0.24

Table 5: Comparison of Original FOM with the difference of mean subjective score

those cases in the 'bad' cubic room as particularly worst than cases in the best Louden ratio room.

Table 5 shows the actual FOM scores, this time with no normalization, and the difference between this and the overall mean score across the three musical samples, again with normalization. The differences between similar FOM cases (there were low,

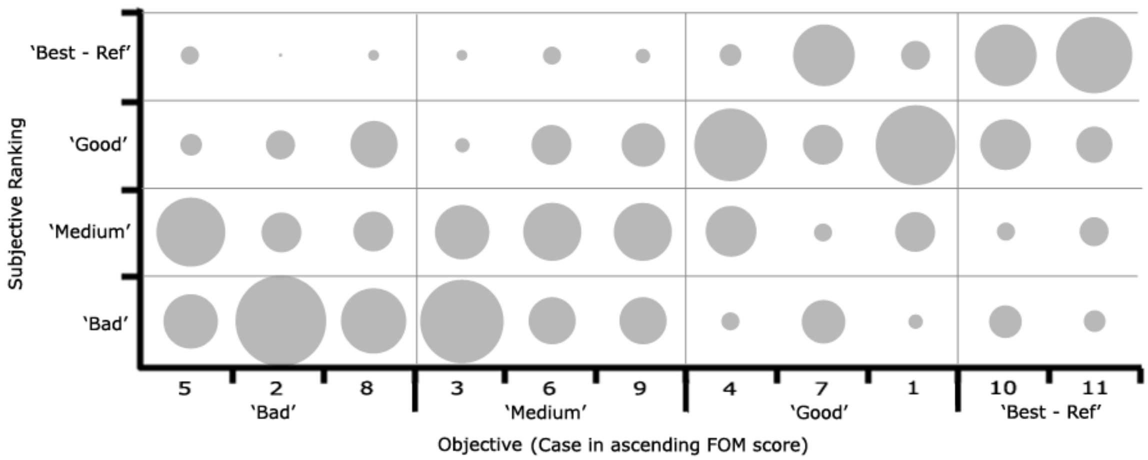

Fig. 6: Scores for each subject categorised into four groups. The size of the circle relates to the number of subjects who ranked the case within the group. For example, a large number of subjects ranked case 2 within their bottom three, ie. 'bad')

medium and high scores for each room) remain reasonably consistent, whatever the room. If the cubic room (room A) was considerably worse sounding at all times, it would be expected that the differences would be significantly greater in these three cases than the other two rooms (both having dimensions according to Louden's best room). Another interesting result is that subjects appear to score 'bad' listening scenarios much more consistently given the much smaller difference between subjective and objective scores for these. This may reveal that although the preference for 'good' listening conditions is somewhat highly subjective, listeners agree much better as to what sounds 'bad'. Perhaps an investigation of what features of a response cause this is worthwhile in defining good room design practice.

5. DISCUSSION

5.1. Reliability of FOMS

Experimental results show that while some trend is visible, correlation results should be seen as indicative only. The high standard deviation suggests that caution should be taken when applying results. It is however suggested that subjects are able to pick out the best room responses according to overall smoothness. However, it is also apparent that at the lower ends of the FOM, where there is a much lower overall smoothness, subjects find differing cases difficult to rate. Some anecdotal notes revealed that in all of these cases the low frequency quality was objectionable. It therefore becomes difficult to rate these samples accurately. It is suggested that where the FOM is low, other quality affecting factors come into play. This is surely to be expected - the peaks and dips in frequency response become more defined and larger in amplitude. It is therefore reasonable to suggest that when this is the case, the interaction of individual frequencies in the musical stimulus with specific peaks and dips in the response is one such key factor. This supposition may also be inferred by the paired comparisons between cases. It was shown that musical sample was, in most cases, insignificant, yet in cases 2, 5 and 9, it was significant. A check of the FOM for these cases reveals that 5 and 2 were low scoring - their frequency responses prone to greater fluctuations in peaks and dips. It would therefore make sense that the musical sample would be more affected by these cases. Care must be taken when making this assessment however, as case 9 scores 0.62 and yet still appears to be prone to musical affects. Interestingly however, an inspection of the frequency response of case 9 does reveal that while the average smoothness may be good, there are two significant peaks and dips at 54Hz and 141Hz respectively. This is shown in Figure 7.

It could be that these frequencies may interact with musical content in some samples to a greater extent than others and cause a decrease in subject rating for this particular room response.

The high standard deviation across subjects must also be considered. It seems there may be differing personal preferences in terms of low frequency acceptability. Further study is necessary to quantify the exact parameters that typical listeners are looking for in 'good bass'. This seems to be somewhat of a grey area and should not be ignored. However, it should be noted that the reference cases, 10 and 11, often scored highly, and were ranked as the two best. It was also noted anecdotally that 'there was always one test case which sounded better'. This was invariably shown to be case 11 - that of an actual smooth response. These comments would indicate that listeners do indeed prefer a smooth response and the characteristics that it brings, such as short decay times and 'punchy bass'. When this is not present, or the characteristics of the room/source/receiver produce an overall difference in amplitude, subjective preference becomes less clear.

The ranking order results seem to suggest that although the exact scores may not be accurately found subjectively, the ranking of cases within categories is in general preserved.

5.2. The effect of musical sample

It is also interesting to note the effect of the three different musical samples. According to the ranking results, sample 2 was the most widely varying. The musical stimulus here was a double bass - a resonant instrument by nature. It is suggested that this increased the complexity of the task for this sample - the resonant characteristics of the source material smearing the ability to judge similar characteristics produced by the room condition. Of all three samples, sample 2 had scores given within the smallest physical range, suggesting similar subjective responses regardless of the room. Conversely, samples 1 and 3 were more alike in tone - punchy bass notes and drum beats. The temporal behavior of elements of these samples should be considered of greater importance. These are affected by the room to a greater extent, and it is suggested that the interactions between low frequency content and frequency response of the room could play a key role in subjective scoring here.

Previous studies on modal spacing and density concluded that the individual interaction of peaks and dips in the room response was of greater importance than statistics derived about the room (modal spacing, density, volume) [5]. Results from this study would agree with this hypothesis, although as expected, when an overall smoothing is achieved, most rooms will benefit. One clear implementation of this is the addition of absorption within the room. This study is also therefore in agreement with previous work which intimates that decay times are the most important factor in subjective modal perception [4]. An investigation into subjective importance of decay times is currently being undertaken. In terms of the lower FOM scoring cases in this test, where interaction may be more significant, a look at case number 5 is revealing. According to the FOM, this case should be the worst tested. The frequency response is the furthest away from a smooth response out of all the cases tested. Figure 4 shows the modelled response and the polynomial curve from which the deviation produces a FOM. According to the FOM, the subjective score given in this case should always be poor. That is, whatever the musical stimulus, the listener perceives a poor quality of low frequency reproduction. However, results show that for music sample 1, the average subjective score was reasonably high, 0.56. In contrast, music samples 2 and 3 both show a lower score. Further study into this relationship between stimulus frequency and room response is necessary, and forms part of the continuing work in this project.

5.3. Significance of Aspect Ratio

The goal of this study of subjective relevance is to determine if these FOMs are valid in use for guidelines in the recommendations of certain room ratios over others. This work would seem to suggest that this method could well have some validity. However, if this is to be followed, caution must be taken. As it is, typical FOM scores across a room do not vary in large amounts. Highest scores, with an increased smoothness, are often rare within the room, unless damping increases drastically in which case the im-

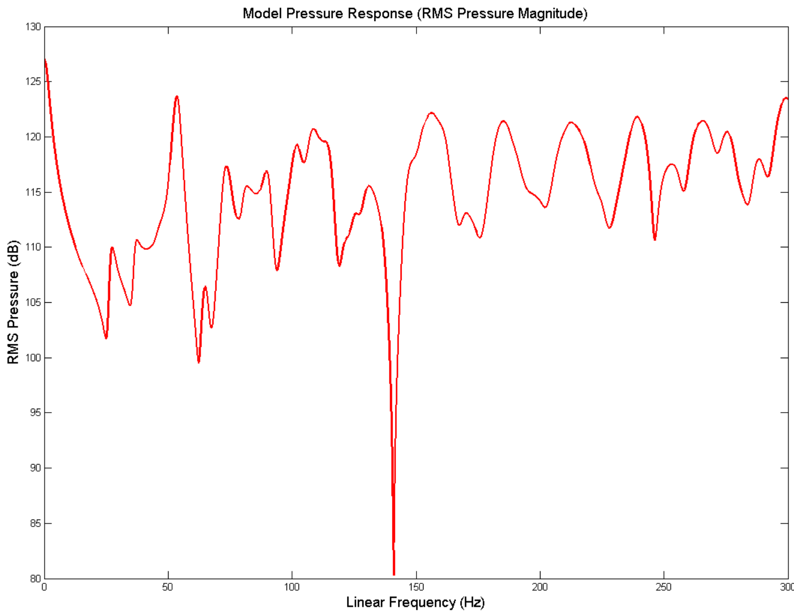

Fig. 7: Frequency Response of Case 9

provement comes from a different control method. Figures 8 and 9 (Appendix 1) show plots across the two aspect ratio rooms. It can be noted that although they differ, the Figures of Merit do not do so by a large amount, and this work has shown that the lower scores, where there are greater deviations in peaks and dips are not shown by subjects to be particularly good in terms of audio quality.

It remains interesting further work to derive averages across rooms of these FOMs in all three dimensions. Such a task however is only relevant if the FOM correlates to subjective perception. This work has shown that this may indeed be the case, (although a greater depth of testing is now required) particularly in the higher FOM region. This kind of study would therefore be of some advantage, especially when considering higher damping rooms. Figure 10 shows the same room as Figure 9 but with higher damping. With a smoother response, shown to be desirable, the average of this room could mean it is capable of sustaining good quality audio throughout. Further study is necessary.

It does remain clear however, that basing room ratios upon modal spacing metrics would appear somewhat flawed. It has been shown that predicted 'bad' rooms, such as cubes, may be scored highly by listeners, and that the FOMs within all rooms, which are based upon smoothness, are roughly similar. Our subjective response would indeed appear to be based closer to this FOM than the modal density or average spacing.

6. CONCLUSION

The reliability of objective measures is an important process in defining any scoring system. This is even more important when using that scoring system as a rational for the construction of listening rooms, often at great expense.

The Figures of Merit have been produced according to the deviation from a smooth fitted response. When a mean subjective score is taken from subjects rating of the low frequency reproduction in modeling rooms, any correlation between predicted score and subjective score can be seen.

It has been shown that in the majority of cases, music sample is not significant. Therefore, when considering an average across all samples, a general trend has been shown. The FOM would appear to be a reliable indicator of good rooms, although its ability to do so is less relevant when the FOM score is lower. The standard deviation of results is high, and it has been identified that subjects may find it difficult to agree on the characteristics of 'good low frequency audio'.

Subjects scores were arranged in the order in which they ranked the 11 cases. These rankings were then broken down into four categories, and it has been shown that there is good agreement between subjective placement and the rank as predicted by the FOM. It is suggested that the FOM can be used with greater effect when used in this way, considering a broader range of catergories, than when attempting an exact match between precise FOM and subjective scores.

It can be seen that the FOM based on the deviation from a smooth fitted curve is of benefit in comparison to the use of metrics calculating a score based upon modal spacing. Such spacing metrics do not account for the source and receiver position within the room. The FOM used can also be linked to the existing damping within the room where other criteria cannot.

Results have also shown correlation with previous research suggesting an importance of individual peaks and dips within the response. Hence, a FOM that looks at the specific response obtained in the room even though it may need adjusting in terms of how weighting for peak or dip deviation and that from a flat line (to account for lack of bass) appears to be a much more robust indicator of room quality and a better basis for design. Obviously, based on this premise, the prescription of room ratios loses importance when compared to optimization of source and receiver position and their interaction to provide a smoother response across an extended area in the room. This is the focus of ongoing work.

7. REFERENCES

[1] F. Everest, The Master Handbook of Acoustics , New York ;London: McGraw-Hill, 2001.

[2] A. Mkivirta, P. Antsalo, M. Karjalainen, and V. Vlimki, "Modal equalization of loudspeaker-room responses at low frequencies," J. Audio Eng. Soc, vol. 51, 2003, pp. 324-343.

[3] T. Welti and A. Devantier, "Low-Frequency Optimization Using Multiple Subwoofers," J. Audio Eng. Soc, vol. 54, 2006, p. 347.

[4] M. Avis, B.M. Fazenda, and W.J. Davies, "Thresholds of detection for changes to the Q factor of low-frequency modes in listening environments," J. Audio Eng. Soc, vol. 55, Aug. 2007, pp. 611-622.

[5] B.M. Fazenda and M.R. Wankling, "Optimal Modal Spacing and Density for Critical Listening," Proc. of the 125th AES Convention, San Francisco: 2005.

[6] R.H. Bolt, "Note on Normal Frequency Statistics for Rectangular Rooms," J. Acoust. Soc. Am., vol. 18, 1946, p. 130.

[7] M.M. Louden, "Dimension-Ratios of Rectangular Rooms with Good Distribution of Eigentones," Acustica, vol. 24, 1971, pp. 101-04.

[8] O.J. Bonello, "A New Criterion for the Distribution of Normal Room Modes," J. Audio Eng. Soc, vol. 19, 1981, pp. 597-606.

[9] R. Walker, "Optimum Dimension Ratios For Studios, Control Rooms and Listening Rooms," BBC Research and Development Report RD1993/8.

[10] C.L.S. Gilford, "The Acoustic Design of Talks Studios And Listening Rooms," J. Audio Eng. Soc, vol. 27, 1979, pp. 17-3.

[11] T.J. Cox, P. D'Antonio, and M.R. Avis, "Room Sizing and Optimization at Low Frequencies," J. Audio Eng. Soc, vol. 52, Jun. 2004, pp. 640-651.

[12] B. Fazenda, M.R. Avis, and W.J. DAVIES, "Perception of Modal Distribution Metrics in Critical Listening Spaces-Dependence on Room Aspect Ratios," J. Audio Eng. Soc, vol. 53, Dec. 2005, pp. 1128-1141.

[13] The MathWorks, MATLAB (R2008b), The MathWorks Inc., Natick, (2008).

8. APPENDIX 1

The following three plots show the variation of FOM across receiver position in two different sized rooms, and with differing damping conditions.

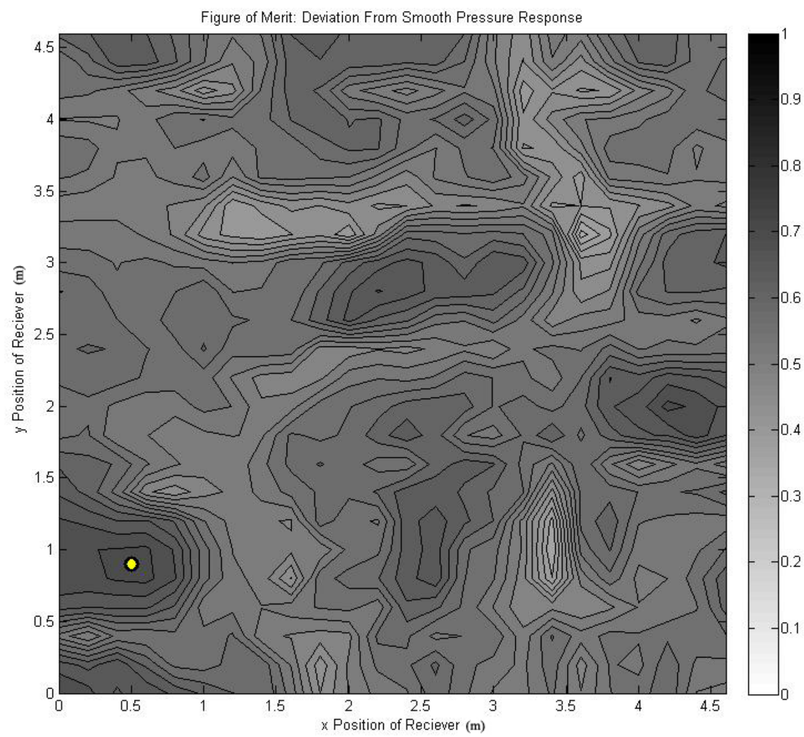

Fig. 8: The cubic room with the damping characteristics used in the subjective test. FOM scores are calculated every 20cm throughout the room. The dot indicates the source position

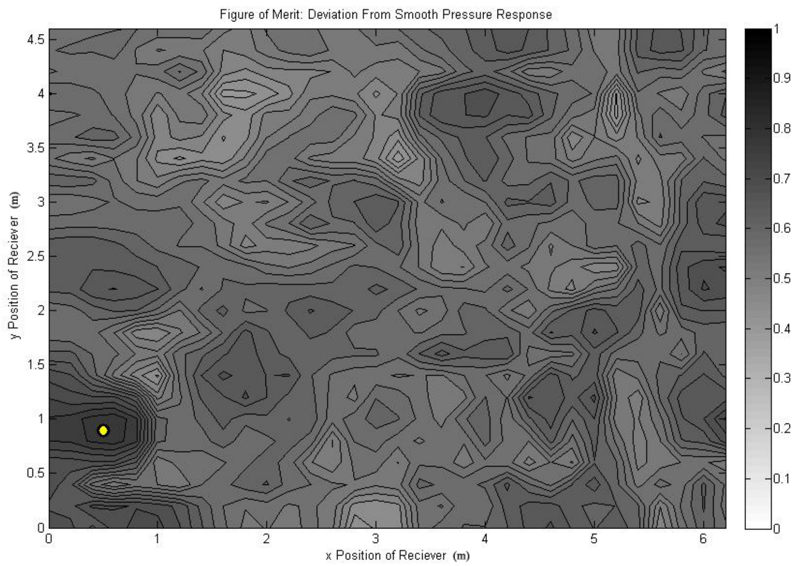

Fig. 9: The best Louden ratio room, here with the damping characteristics used in the subjective test. FOM scores are calculated every 20cm throughout the room. The dot indicates the source position

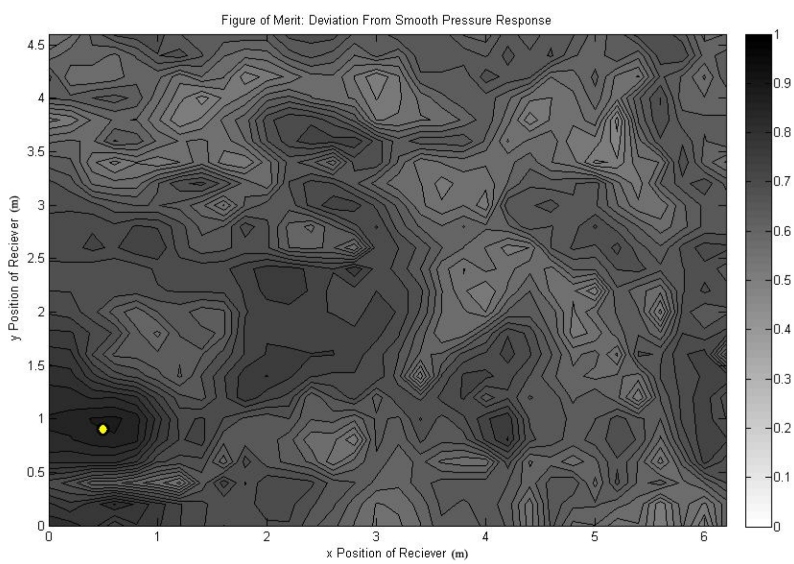

Fig. 10: The best Louden ratio room, here with a higher damping. Results suggest that a smoother response, as produced by higher damping is indeed subjectively relevant. Therefore the greater uniformity of this room may well produced good all round audio reproduction quality