Audio Engineering Society

Convention Paper 7584

Presented at the 125th Convention
2008 October 2–5 San Francisco, CA, USA

Optimal Modal Spacing and Density for Critical Listening

Bruno Fazenda[1], and Matthew Wankling[1]

[1] University of Huddersfield, Huddersfield, West Yorkshire, HD1 3DH, England
b.fazenda@hud.ac.uk, m.wankling@hud.ac.uk

ABSTRACT

This paper presents a study on the subjective effects of modal spacing and density. These are measures often used as indicators to define particular aspect ratios and source positions to avoid low frequency reproduction problems in rooms. These indicators imply a given modal spacing leading to a supposedly less problematic response for the listener. An investigation into this topic shows that subjects can identify an optimal spacing between two resonances associated with a reduction of the overall decay. Further work to define a subjective counterpart to the Schroeder Frequency has revealed that an increase in density may not always lead to an improvement, as interaction between mode-shapes results in serious degradation of the stimulus, which is detectable by listeners.

1. INTRODUCTION

The problem of resonant modes in listening spaces has long been acknowledged. Reducing the negative perceptual effects of these modes is fundamental for room designers aiming for the highest quality of audio reproduction and to loudspeaker manufacturers aware that this is one aspect that can severely affect the perceived quality of their product. Due to the relationship of these modes with the physical dimensions of the room, researchers have often looked at optimal room aspect ratios in an attempt to avoid modal degeneracy – multiple modes overlapping at the same frequency. Work of this nature has often concentrated on attempts to control the distribution of all possible modes in a given room [1,2]. More recently, the particular response dependent on source and receiver position has been acknowledged as more representative of the general use of such rooms [3,4]. In any case, the frequency spacing between adjacent modes and their density in a given frequency range has been fundamental for all studies of the low frequency modal behavior of these spaces. This paper studies the perception of these two related areas, modal spacing and modal density.

2. MODAL THEORY

Modal spacing and density have often been used as objective measures to quantify the quality of reproduction in a listening space. Modal spacing theory has suggested that an increase in room acoustic quality is associated with a greater uniformity of spacing in frequency between adjacent modes. Optimal room ratios such as those published by Louden [1] attempt to optimize this spacing. More recent work by both Cox [3] and Fazenda [4] has also focused on the subject of optimal room ratios and considered objective metrics by which it may be possible to classify the room response.

When considering the effects of modal distribution on the sound quality of a room, it is generally accepted that a flat frequency response is desirable. The presence of peaks and dips modify the overall sound for the listener by altering the amplitude at certain frequencies. Furthermore, the Q-factors of these peaks and dips are also associated with decay times for a particular frequency. In comparison, the flattest response, corresponding to a lower Q-factor, results in the shortest decay time and in general the more homogeneous frequency responses (flat) are associated with shorter time responses. It follows that an arrangement of the modal frequencies corresponding to a more homogeneous frequency response will result in shorter decay times in the modal region and consequently to an improvement of the audio reproduction quality. This paper examines whether an optimum spacing between resonances can be defined which is associated with the shortest decay time of the system and hence the best perceptual condition. If available, this metric could in turn be incorporated into room design at low frequencies. Objective measures such as the Modulation Transfer Function (MTF) are presented, and conclusions drawn to their relevance in relation to subjective results. This is described in Section 3.

Further objective measures have considered the modal density. Examples include the 'Bonello Criterion' [2] and the widely quoted 'Schroeder Frequency', which defines a transition frequency between the 'modal' and 'statistical' sound-fields [5] in a given room. This transition frequency is determined by equation 1.

$$f_c = 2000\sqrt{\frac{\mathrm{T}}{\mathrm{V}}} \qquad (1)$$

where f_c is the transition frequency, T the 60dB reverberation time in seconds and V the room volume in m^3.

This value identifies the frequency above which at least three modes fall within one bandwidth of one mode. In some cases it is implied that above this frequency, within the 'diffuse' region of the sound-field, the individual effects of resonances are no longer perceived. Many research papers use this somewhat arbitrary value as a limiting point for their investigations into the effects of low frequency resonances. The work of Avis et al. which investigates the perception of room modes, uses the Schroeder Frequency as the point of transition when forming binaural room models [6]. In their 'Room Sizing and Optimization' paper, Cox et al. also state that the frequency range under investigation can be "guided by the Schroeder frequency" [3]. Furthermore, Toole states the importance of the crossover region as a real phenomenon which needs to be better understood [7].

As the size of an enclosure reduces, the Schroeder Frequency rises. In large rooms such as concert halls, this frequency is typically very low, often below the 20Hz threshold of our hearing. However, spaces such as control rooms, of typically small volume (i.e. $100m^3$) are affected by the modal sound-field at frequencies not only above 20Hz, but well into the range of most musical situations (i.e. T=1.28s, $V=75m^3$, f_c=261Hz – middle C). This becomes a problem as the modes then have the potential to degrade the original musical signal. We must therefore, seek to gain a better understanding of the subjective nature of this transition region.

Section 4 of this paper presents the results of initial work towards a subjective counterpart to the Schroeder Frequency, supporting a better understanding of where our perception of audio quality is no longer related directly to measurable modal parameters.

3. MODAL SPACING

Theoretically it is possible to define an optimal spacing between two adjacent resonances which results in the shortest decay time of the whole system. It is hypothesized here that a subjectively optimal modal spacing also exists and can be measured.

3.1. Objective Measures

3.1.1. Visual Examination

Figure 1 represents the response of a system comprised of two resonances. A simple visual investigation of the effect of altering the spacing between the two individual resonances reveals a clear reduction in decay time. However, as the second frequency moves away from the first, the magnitude frequency response reveals a large dip and the resulting impulse response begins to show a distinctive amplitude modulation. This is obviously associated with the interaction between the two resonances and at these frequency differences they sound identical to 1st order beats as described in many psychoacoustic textbooks [8]. When plotted as a logarithmic decay (Figure 2) the beating effects are even clearer.

One can make assumptions based upon this visual inspection as to the perceived quality of an audio stimulus when passed through these resonant systems (assuming the audio material were to excite the corresponding frequency range). The shortest decay is clearly preferable, while the introduction of beats will be highly detectable to the listener and perhaps undesirable. The question however remains; at what point along this sliding spacing scale does the optimal compromise between the two degrading effects lie?

Without such a simplified system of two carefully spaced modes of identical amplitude and phase, a simple visual examination of the time domain response becomes increasingly difficult. Thus, a computational method for predicting the same result is desirable.

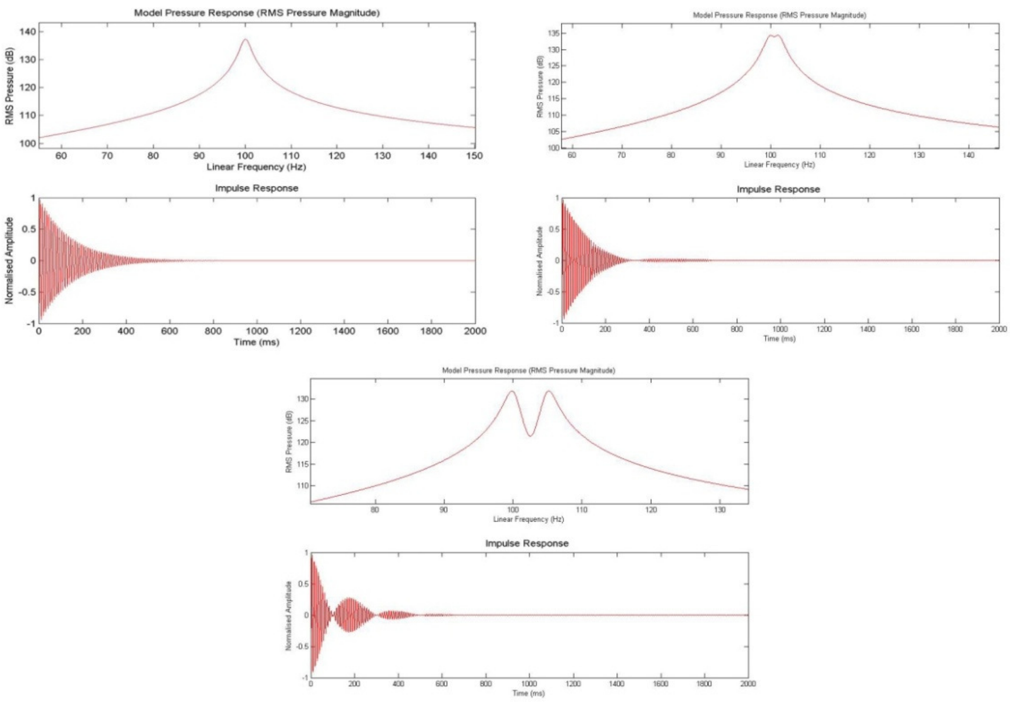


Figure 1: a) 100Hz & 100.1Hz b) 100Hz & 101.5Hz c) 100Hz & 105Hz

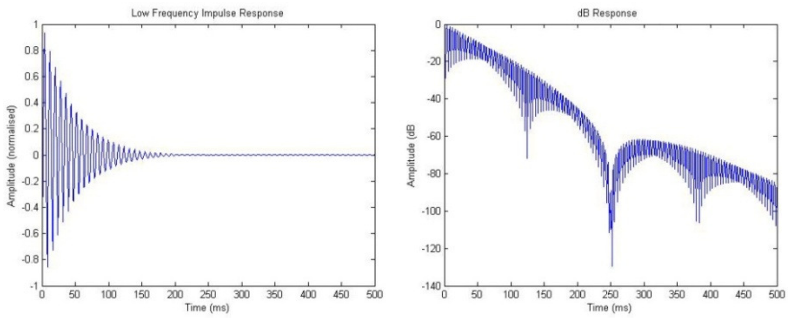


Figure 2: The computed response displayed as a normalized impulse and also in dB

3.1.2. The Modulation Transfer Function

The Modulation Transfer Function (MTF), originally developed in the field of optics as a quantifier of lens image resolution, has also been shown to correlate well with audio reproduction quality [9-11]. It measures the system's ability to preserve amplitude modulations of a signal over a set frequency range. The modulation frequencies are defined as representative of audio signals and in particular those found in speech where this technique is applied to define a speech transmission index. The function takes the input response of the system and calculates a figure of merit between 0 and 1 with the top of the scale corresponding to an exact copy of the input signal.

Resonances were generated using the Green's Function (Equation 2) which has previously been used to successfully model low frequency room responses [3,4,12]. A fixed array of the two modal frequencies was fed into the decomposition equation to obtain the system's response. These impulse responses were then passed through the MTF algorithm (see [11]), which was adjusted to determine the result in the frequency range of the modes.

$$P_{\omega}(r) = j\omega\rho Q c^2 \sum_n \frac{P_n(r)P_n(r_0)}{X_n(\omega^2 - \omega^2{}_n - 2j\delta_n\omega_n)} \qquad (2)$$

Variables under test were the frequency range of the modes and the Q-factor. As the Q increases, the resonant peaks become sharper and a greater definition between individual frequencies is detectable. Measurements were carried out at three test frequencies, 63, 125 and 250Hz. Figure 3 shows an example of the MTF mapping across a range of modal spacing for a number of modal Q-factor values. The modal frequency in the example is 63Hz.

It is clear that the MTF results indicate the same trend evident in Figure 1. For a given modal Q, there is an optimal modal spacing associated with a peak in the MTF score (around 4Hz in Figure 3). It is interesting to note that as spacing continues to increase, a number of local minima and maxima are predicted by the MTF. As expected, a reduction of the Q-factor increases the predicted optimal spacing. However, it is clear that at these low Q values the score is largely independent of modal spacing. It is interesting to confirm that MTF predictions in this case are in line with previous findings that suggest low Q modes to be less problematic[see for example 6].

Table 1 shows the optimal spacing as predicted by the MTF metric at each frequency and for increasing values of Q-factor.

Frequency (Hz)	Q=10	Q=20	Q=30	Q=40	Q=50
63	8.5	5.3	4.1	3.5	3.3
125	12.6	8.4	6.5	5.3	4.6
250	21.6	12.6	9.9	8.4	7.4

Table 1: Optimal Spacing as Predicted by MTF (Hz)

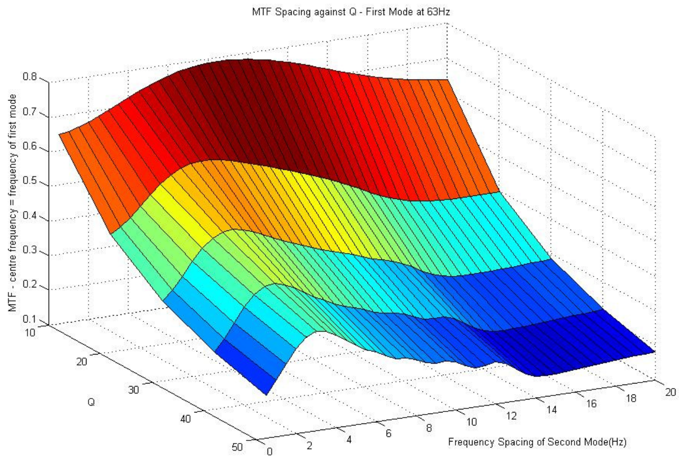


Figure 3: Example of MTF scores across spacing at different Q's - frequency of first resonance 63Hz

3.2. Subjective Test

For the subjective tests, the two spaced resonances were artificially modeled using the same method described above. The resulting frequency response was transformed to the time domain, giving the impulse response of the 'room' in question. Whilst this impulse could be convolved with an input stimulus such as a test tone or musical refrain, it was decided that the impulse itself should be used as the test stimulus since its effects are distinct and more audible than using any other input stimuli. Single frequency decaying sine tones were considered, but the decay length of the tone would in some cases be responsible for masking the decay of the resonance itself. As such, a threshold measurement corresponding to 'the worst case scenario' was found to be adequate.

The same three resonant frequencies with four Q factors (10, 20, 30, and 50) were chosen to represent a broad range typical in listening conditions. The spacing of the second resonance was adjusted by way of a slider on a graphical user interface (Figure 4). Samples were generated instantly each time the slider was moved, removing resolution error from pre-defined steps. All programming was carried out in Matlab. During each test, subjects were asked to adjust the spacing slider to the point where the overall decay sounded the shortest. Prior to the test, explanation of the differences in presentation sounds (long decay, shorter decay, and beating effect) were explained, along with images in the time domain. It was also explained that beats were to be considered as part of the overall decay process. No time domain images were displayed during the actual tests to avoid bias.

Eleven subjects were tested, in quiet studio conditions, with samples auditioned over a pair of Sennheiser HD-650 headphones. Each subject was given time to practice before the test commenced. The presentation levels of the three frequencies were weighted to ensure that the perceived level of each sample was the same - samples were presented according to the 90dB equal loudness contour [13].

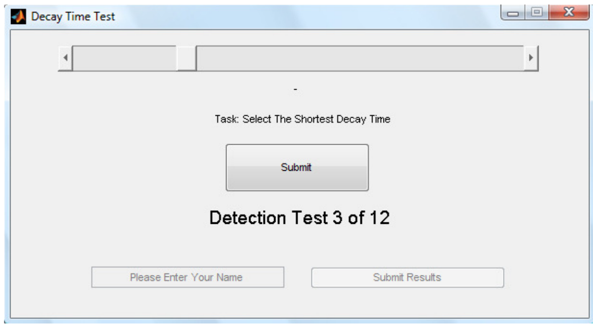


Figure 4: Screenshot of Spacing Test

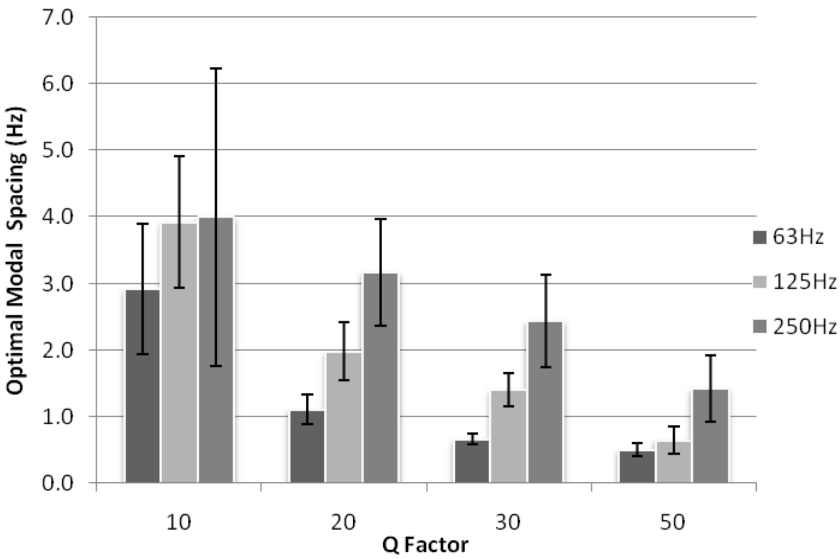


Figure 5: Mean Spacing across Q Factor and Frequency

3.3. Results and Analysis

Results are shown and statistical analysis has been carried out to show the significance of each result.

Figure 5 shows the mean spacing across 11 subjects. A simple visual inspection reveals clear trends. As the Q factor increases, the optimal spacing needed to provide the shortest decay reduces, as expected. When comparing the test frequencies, again it is clear that higher frequencies require a greater spacing between the two resonances. It should be noted here that this is in direct contradiction to the natural decrease of modal spacing in rooms as frequency increases. Furthermore, the level of uncertainty, shown by the standard deviation error bars also increases with frequency indicating that an optimal spacing becomes less meaningful as frequency increases.

Analysis of variance was carried out to ascertain the level of significance across the variable parameters. Table 2 shows that both the Q Factor and modal frequency are highly significant, i.e. $p<0.01$, which indicates the success of systematic testing.

Experimental Factor	***p***
Q	0.00
Frequency	0.00

Table 2: Anova Test

Although both factors are highly significant, it is useful at this point to wrap them into a single factor - that of modal bandwidth. Frequency, Q and bandwidth are related according to the equation:

$$Bw = \frac{f}{Q} \tag{3}$$

Table 3 considers each of the 12 test scenarios in ascending bandwidth. The results again show a clear trend:

BW	1.26	2.10	2.50	3.15	4.17	5.00	6.25	6.30	8.33	12.50	12.50	25.00
Q	50	30	50	20	30	50	20	10	30	20	10	10
Freq	63	63	125	63	125	250	125	63	250	250	125	250
Mean	0.5036	0.6643	0.6458	1.1079	1.4075	1.4284	1.9860	2.9183	2.4411	3.1664	3.9237	4.0013
St.Dev	0.0959	0.0866	0.1998	0.2220	0.2512	0.5007	0.4355	0.9729	0.6961	0.8013	0.9843	2.2361

Table 3: Mean Subjective Optimal Spacing presented in ascending Bandwidth

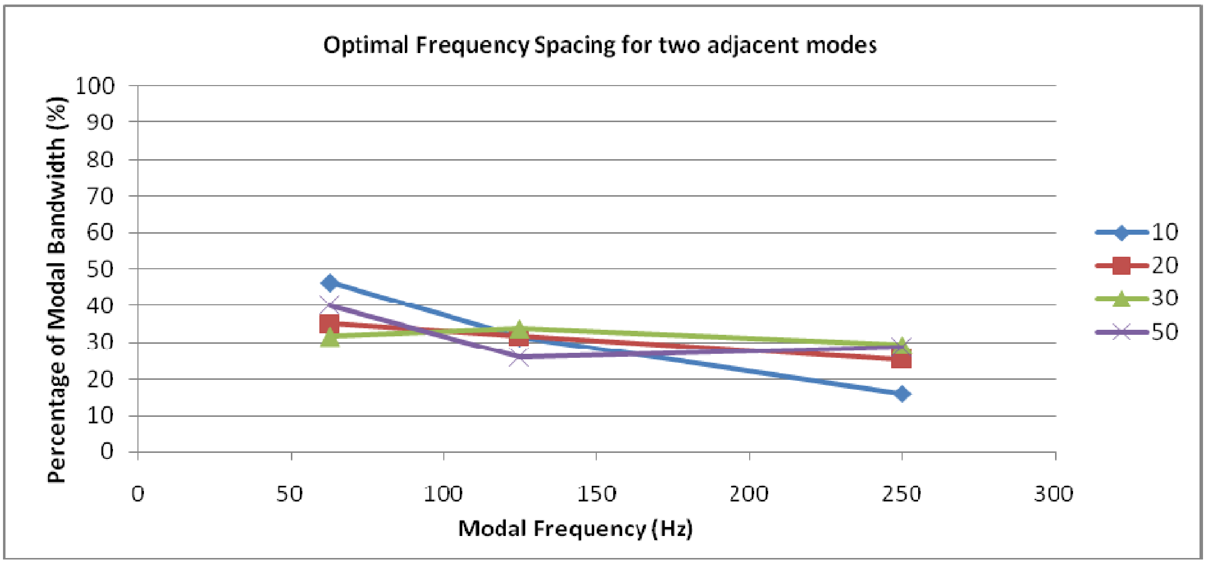


Figure 6: Optimal Spacing across ascending bandwidth for the four different Q Factors tested

Figure 6 shows optimal spacing as a percentage of the modal bandwidth. This figure reveals that, for Q's of 20, 30 and 50, regardless of frequency or Q, the optimal spacing lies between 25 and 40%. At lower Q's, the standard deviation becomes higher (see Table 3) and results are less reliable. These results were confirmed by comments from subjects who each stated that the shortest impulses were significantly harder to judge than those of longer length.

3.4. Discussion

The results relevant to the subjective perception of an optimal modal spacing are now discussed. In this investigation, it is clear that, when using a simplified scenario of two single resonances, the decay time imposed by the response of the system can be optimized by an ideal spacing of their centre frequencies. As the bandwidth of each resonance increases, so does the optimal spacing. As the two frequencies separate further, a dip in the response can be identified, which in turn leads away from a flat shape, and beating between the two frequencies becomes identifiable.

Results are encouraging in defining a trend. However, there are a number of points to note. Firstly, although clear results have been identified, further investigation would suggest that the listening level may have a significant impact. It is possible to relate the spacing values obtained to the point where a first beat occurs at a level of -60dB relative to peak loudness of sample. Table 4 shows a correlation between the measured values and the peak level of the first beat. As it should be expected, with louder listening levels, the beat peak amplitude becomes louder, and there is some evidence from subsequent testing by the authors that the spacing would reduce (as the beat is heard sooner).

Comparison between subjective test results and those predicted by the MTF, reveals that although they differ significantly in value, the same trend is clearly apparent – an increasing in optimal spacing with increasing bandwidth. Therefore it would seem that an adjustment of the MTF metric, or indeed, a metric with better correlation to perception could accurately predict the subjective optimal spacing between the two resonances.

The subjective results reveal that at these low frequencies, a much closer spacing is needed than is usually achieved by room design. Also apparent is the fact that the effects of poor modal spacing are more noticeable at the lower range of those frequencies studied, giving weight to the argument that it is at these lowest frequencies that modal optimization should be focused. At 250Hz, the differences in spacing were very difficult to perceive. Furthermore, at the lowest tested Q value of 10, spacing differences were also difficult to perceive. This result is in agreement with previous research which suggests a threshold for detection of changes in modal Q-factor at around Q=16 [6].

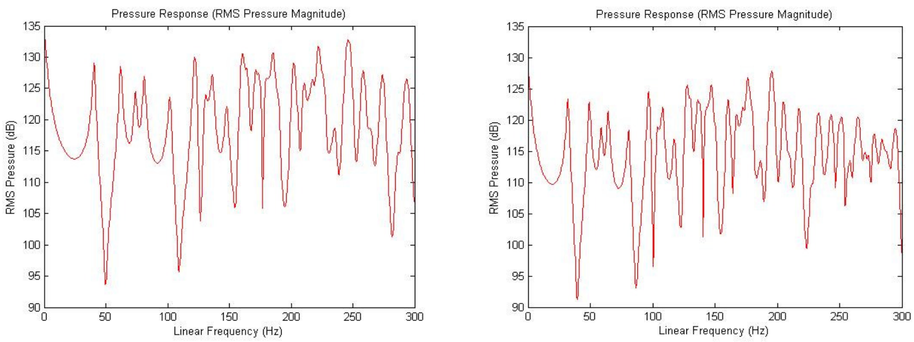
Finally, these results open up further research avenues. For example, will the masking effects of a musical stimulus cause a difference in result, or will the same detection of the shortest decay and onset of beats remain? Further work currently being undertaken also looks at the effects of multiple modes rather than the simple pair used in this test.

4. MODAL DENSITY

As stated, modal spacing decreases with frequency in rooms. Therefore modal density increases. Eventually many hundreds of modes lying within a few Hertz exist. It is this increase in modal density that underpins the definition of the Schroeder Frequency as a transition region from 'modal' to 'diffuse' sound field. Another aspect that influences an increase in modal density is the volume of the room – larger rooms have a higher modal density than smaller rooms for a given frequency range. Moreover, if the aspect ratio of the room remains constant, as volume increases, the modal frequency response retains the same shape, only 'squashed' into a narrower frequency band (Figure 7).

BW	1.26	2.10	2.50	3.15	4.17	5.00	6.25	6.30	8.33	12.50	12.50	25.00
Mean	0.50	0.66	0.65	1.11	1.41	1.43	1.99	2.92	2.44	3.17	3.92	4.00
-60dB	0.42	0.69	0.83	1.04	1.38	1.66	2.07	2.07	2.76	4.12	4.12	8.20

Table 4: Subjective optimal spacing compared with the calculated spacing at the point where the first beat amplitude at -60dB



a) 50m3 b) 100m3

Figure 7: 'Squashing' of Frequency Response as room volume increases

It is assumed that as a large number of modes are concentrated in a given frequency range, as would happen with a volume increase, the overall magnitude frequency response becomes 'flatter' and thus is commonly associated with better quality reproduction. This section tests the subjective relevance of this argument.

4.1. Test omitting the Mode Shapes

The Greens Function (Eq. 2) for modal decomposition is once again used to generate room responses. Subjects were asked to increase the volume of a sample room until there was no perceived difference when comparing with a smooth (flat) response containing a reference density. This then identifies the detection threshold where the modal density of the variable room is perceptually the same as that of the reference. The density at a given frequency can then be extrapolated using an expression describing typical mode spacing in rectangular rooms [14].

During pilot testing, it became clear that such a threshold was achieved only if the mode-shapes (Pn(r) and Pn(r0) - the coupling of source and receiver positions in equation 2) were omitted from the model.

Although somewhat unrealistic, this condition replicates the case where all modes are simultaneously excited and received, which represents the conditions assumed for room ratio metrics as suggested by Louden, Bonello, Bolt etc [1,2,14]. In practice, these conditions are never actually attained in rooms but they can be considered as the case of the ultimate 'smooth' response in modal terms. This target could be used in low frequency diffusion design or in correction techniques that artificially add modes to smooth out the existing response at a given position in the room – although for all cases modes need to add constructively.

A set of tests was run omitting the mode shapes in the model (by setting Pn(r) and Pn(r0) both equal to 1). In this case the response flattens out as density increases (see Figure 10b). PEST (Parameter Estimation by Sequential Testing) methodology [15,16] was employed to home in on the subject's threshold of detection between a reference sample, in a room of 100000m3, and that of a second sample within a room of a variable volume. To ensure that the subject could not simply claim to hear a difference, an ABX procedure was employed. At each volume three comparisons were made. If the samples were correctly identified three times in a row, the volume is increased. However, a single incorrect answer would immediately register a failure to detect a difference and therefore the volume would decrease. The requirement of three consecutive correct answers reduces the probability of the subject guessing to 12.5%, and while this is not at the typical statistical threshold (<5%), it was considered sufficient given the association with the PEST methodology, which would bring the volume back down at the next comparison unless six consecutive guesses were made - a probability of just 1.6%.

Test tones (0.4 second decaying sines) were used at the same three octave bands as in the spacing test, 63Hz, 125Hz and 250Hz. These tones were convolved with the modeled room response. Once again samples were weighted and presented according to the 90dB equal loudness contour. Eight subjects were tested, under the same conditions as for the spacing test.

Figure 8 shows results for the mean value and standard deviation for room volumes where no detectable difference existed between the two cases compared.

In practice the results provide the preferred density for a particular frequency. However, to extract the modal density at these three cases, a modal bandwidth for the corresponding frequency has to be obtained from the damping conditions in the model (δ). Modal density can then be calculated as the number of eigenfrequencies within a modal bandwidth. This can be achieved using Bolt's equation as follows:

$$\frac{\delta N}{BW_{modal}} = \frac{4\pi F^2 V}{c^3} \quad (4)$$

where F is frequency, V is room volume.

This density is indicated in Table 5.

Frequency (Hz)	63	125	250
Modal Bandwidth as prescribed in the model - (2.2/RT)	2.17	2.63	3.75
Subjective Volume Threshold	1529	803	433
Subjective modal density (Eq. 4)	4.1	10.3	31.6

Table 5: Modal Density According to Bandwidth from model damping conditions and subjective volume threshold

The results show that at 63Hz a subject would require around four modes per modal bandwidth to even out modal effects. Schroeder's theory requires three or more modes to prescribe a diffuse sound field. Furthermore, under these test conditions, subjects require an increasing modal density as frequency rises. This is shown in Table 5 where a volume associated with a larger density is selected as the threshold. Consequently, no definition of a generic modal density across frequency is possible from these results. Although at the very low frequencies a modal density of about four is sufficient and in accordance with the definition for the Schroeder Frequency, as frequency increases subjects prefer even more modes together.

In itself this is an interesting result. However, as discussed previously, any realistic scenario should include the effects of the mode-shapes as these carry crucial information about the way in which the source and receiver position couple with the modes.

4.2. Mode Shapes

An alternative and more realistic scenario is when the mode shapes are included. In this case, Pn(r) and Pn(r0) take relevant values related to source and receiver positions giving a somewhat different response (Figure 9a).

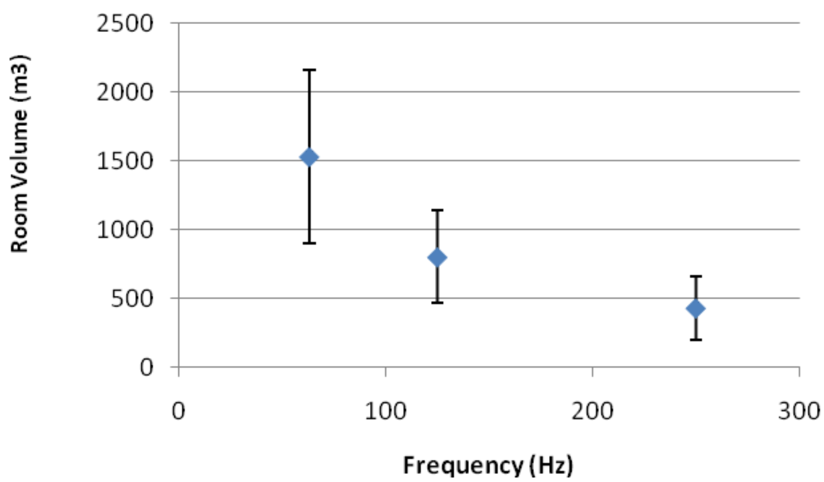


Figure 8: Mean threshold volume for the detection of difference over three test frequencies

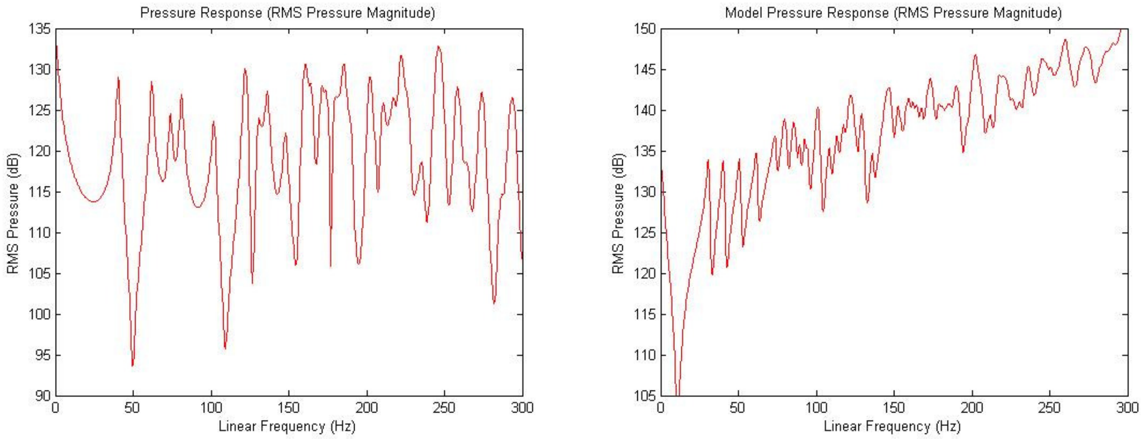


Figure 9: a) with mode shapes, b) without mode shapes (room volume 50m3)

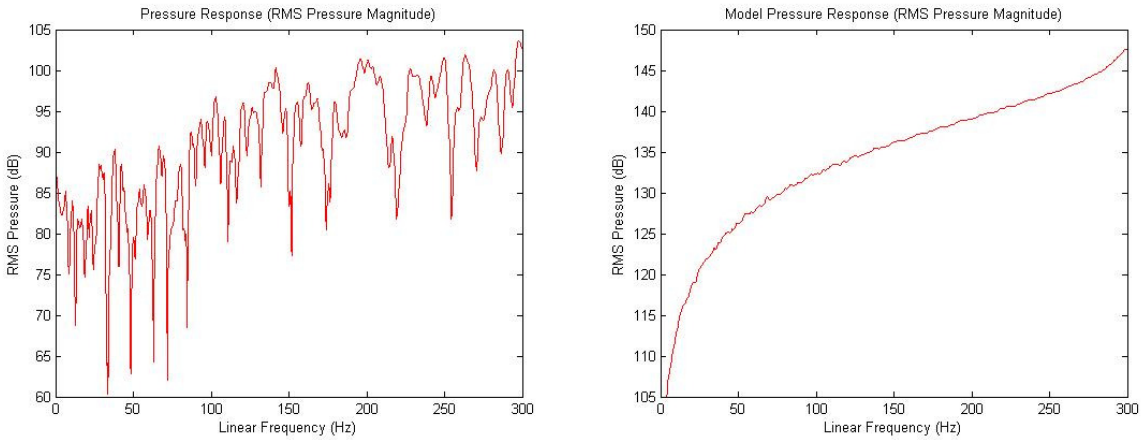


Figure 10: a) with mode shapes, b) without mode shapes (room volume 10000m3)

For higher room volumes, the difference between the two approaches is striking (see Figure 10). It is clear the two responses are not the same! The differences of course arise due to the interaction between the modes. At this volume, a bandwidth of just 1Hz at 125Hz already contains around 60 modes corresponding to a modal spacing of 0.017Hz. During pilot testing for the same density threshold as in the case with no mode-shapes, convergence was never achieved. Subjects were able to detect differences even at unrealistically high modal densities. In order to test the effects of density increase including the effects of modal coupling a more robust approach was needed.

A further test set out to study how accurately listeners detect differences in modal density when the mode shapes were included and took relevant values related to the source and receiver positions. To test this, a simple ABX test was conducted, consisting of ten paired comparisons. It had already been noted that with test tones a difference is always perceptible. Hence, to increase the realism of the test a musical stimulus was chosen. Sample A was a reference room modeled at a specified volume. Two reference volumes were tested - 500m3 and 10000m3. Sample B varied in volume approaching the reference. Sample X was the unknown sample that the subject was asked to identify as A or B.

Each of the ten ABX tests was fixed at 10 trials. The same eight subjects were tested as with no mode-shapes. Results are presented in Table 6 and Figure 11. In addition to the actual volume of the target room, the volume is indicated as a percentage to enable comparison between the two cases tested.

The same trends are evident for both room sets. Regardless of general volume, if the compared rooms are very different, detection is a simple task. This task remains relatively simple until the differences in volume are below 10%. At this point, the frequency response is very similar and detection is no longer possible.

A chi-square test was carried out on the data to determine the significance of each result. Values for p indicate the success of detection in each case. Values below 0.05 report a significant detection whilst above this value no detection is validated. Therefore, the statistical results show the same trend for both room sets – large and small. It becomes increasingly difficult to detect a difference as the volume approaches that of the reference room. Above around 90%, the subjects are not able to tell the difference significantly.

The interesting outcome is that even in large rooms, where modal density is inherently high, there is no significant reduction of audibility of modal effects. If, as the Schroder Frequency theory suggests, the sound field becomes more diffuse, then these results do not suggest that our perception follow those of diffuse conditions.

Small Room	Reference Volume	500	500	500	500	500
	Test Room Volume	100	250	400	450	490
	% of reference	20%	50%	80%	90%	98%
	Mean correct identifications	9.22	8.56	8.33	8.11	6.56
	p	0.0000	0.0011	0.0042	0.0057	0.1512
Large Room Volume	Reference Volume	10000	10000	10000	10000	10000
	Test Room Volume	1000	5000	9000	9500	9990
	% of reference	10%	50%	90%	95%	99%
	Mean correct identifications	9.11	8.56	7.67	5.89	5.89
	p	0.0001	0.0008	0.0244	0.1342	0.9212

Table 6: Results and Chi-Square analysis showing the mean correct identifications and significance of each test - $p<0.05$ indicates the subjects could significantly identify different rooms. Percentages refer to the percentage volume of the test room (sample A) compared to the reference (sample B).

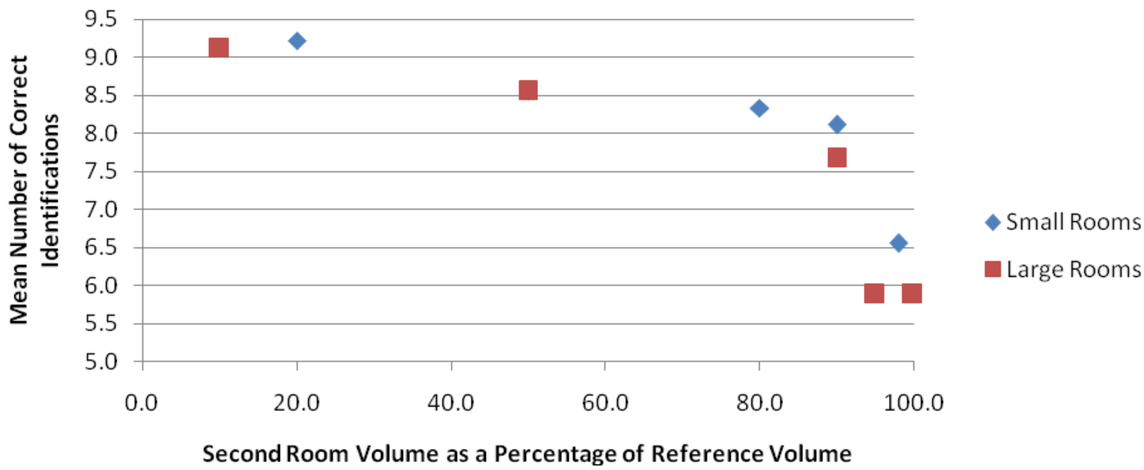


Figure 11: Correct Answers in the identification of Two Room Volumes

5. CONCLUSION

5.1. Optimal Modal Spacing

A subjectively defined optimal modal spacing has been measured. This metric is shown to increase with frequency and decrease with Q-factor. When specified in terms of percentage of modal bandwidth, the optimal spacing lies between 25% and 40% of modal bandwidth regardless of frequency and Q (with exception to a Q value of 10).

The reliability of subjects responses also show that modal spacing is important at the lowest modes but its significance decreases with increasing frequency. A smaller spacing than optimal leads to longer but homogenous resonant decays. This has been shown to be problematic for sound reproduction [4,7]. However, larger spacing than optimal leads to beats in the decay. The relative importance of these two factors (long single decays vs. perception of beats) has not been measured and it stands out as an interesting avenue for future research. It should be noted that this applies mainly to case where two resonances share a very narrow band of frequencies which is representative of the lowest modes in a given room.

The measured results were compared to predictions from an objective measurement – the MTF. Comparison reveals that the MTF may predict trends in room performance, although in its current state it does not match the subjective responses identified here. Refinements to the metric may well achieve this in the future.

5.2. Optimal Modal Density

Tests concentrating on more realistic room scenarios focused on the definition of an optimal modal density.

A condition where the effects of source and receiver coupling to the mode-shapes are omitted has been used to study the required modal density that evens out the frequency response satisfactorily. Results from this study reveal that there is indeed a convergence where listeners can no longer perceive differences between two rooms of differing volumes and hence of differing densities. This would suggest that an optimal modal density has been reached.

At the lower range of frequencies tested, around four modes per modal bandwidth are necessary. This number should then increase with frequency and at the higher range, 32 modes per bandwidth are required. This, to some extent, contradicts the general belief that modal degeneracy is problematic. Indeed, a number of modes all sharing the same very narrow frequency band is unwanted, and this is clear from the results shown in the optimal spacing case presented. However, as modal density increases with room volume or frequency, many cases of modal degeneracy exist in the responses that are not perceived as being problematic.

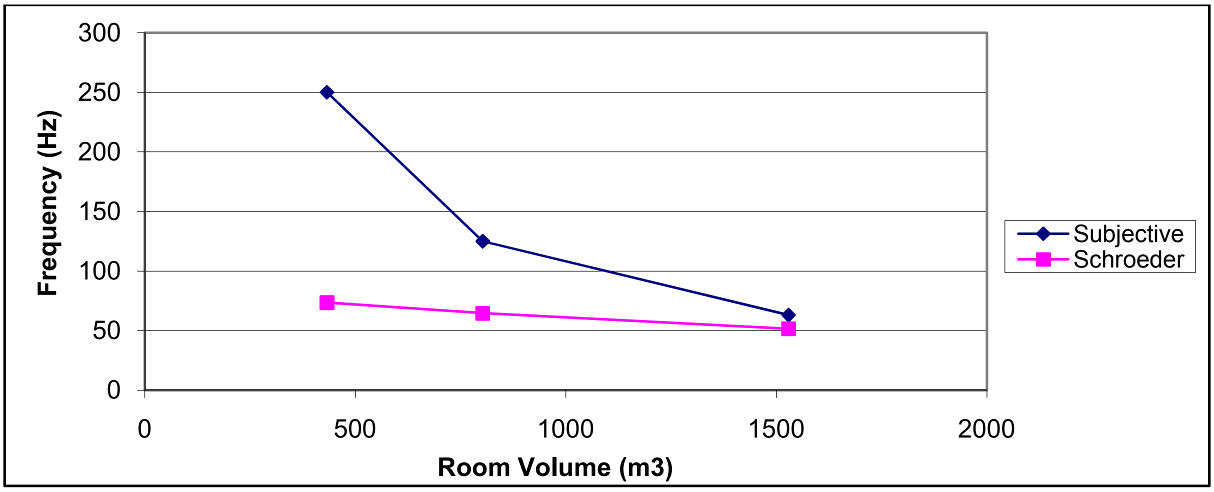


Figure 12: 'Cut on' Frequency for 'diffuse' conditions

Another way of reading these results is shown in Fig 12. The subjective 'cut-on' frequency above which modal effects are negligible is indicated both from these subjective tests and determined from the Schroeder Frequency (Eq. 1). It is clear that for small rooms the Schroeder Frequency underestimates the subjective 'cut-on' frequency – subjects still detect differences in modal sound fields above Fs. For larger volume rooms, the subjective results converge to Fs.

For tests where mode coupling is accounted for, this theory breaks down. No single point was found above which modal density becomes high enough to produce a response which sounds subjectively the same as a reference.

The same trend is seen for both typically large and small rooms. The large rooms tested here have a much higher modal density than the small rooms, and yet the same results are observed – subjects can reliably detect a difference between modal sound-field until the room volumes differ by less than 10%, at which point reliable detection is no longer possible. It appears that detection of differences in modal sound-fields is strongly influenced by the mode-shapes. Hence, one cannot dismiss the actual effects of the response solely on the basis of modal density. These results suggest that it is the interaction of modes with the source/receiver position that determines the perceived audio quality. During pilot tests, anecdotal evidence from a number of listeners suggested that there was no continual improvement in the reproduction quality as the density was increased, rather, there were sporadic points across a range which sounded better than others. Initial investigations into this would seem to suggest that dips in the frequency response are responsible for lower audio quality. This is to be the subject of further research.

5.3. Final Remarks

In conclusion, the results from these studies raise some interesting issues.

It is clear that modal optimization processes that attempt to relocate modal frequencies by changing room dimensions must take into account the coupling of source and receiver positions in the room. Indeed, this necessarily becomes another optimization variable as explored by Cox et al. amongst others [3].

At the very low frequencies, modal degeneracy is certainly problematic. Its effects are long resonant decays if modes are too close in frequency and amplitude modulation beats if too far apart. In this region, where modes are sparse and modal control is more challenging, an approach to space the modes optimally is worthwhile. Optimal spacing of about 25% to 40% of bandwidth as indicated in this study can be used as a guide. The prescription of aspect ratios, source/receiver positions and low frequency diffusion methods are all useful to achieve this.

At higher frequencies (>125Hz), where density increases, the interaction between modes is such that modal effects are still noticeable regardless of density. At these frequencies, the interaction of stimuli and particular room response at its frequency is once again proven crucial - see Fazenda et al. [4] for another example. The concept of high modal density is not directly linked to improved perception.

The 'resonant' characteristic of modal sound is certainly associated with low modal density, as in these conditions, most of the excitation signal is concentrated on the modal frequencies especially during the natural response of the room. This is indeed what is commonly perceived as the difference between modal and 'diffuse' sound-fields. In this case, an increase in modal density is helpful if it fills the frequency 'gaps' between the modes, resulting in a more homogeneous decay across frequency. However, if the decays are still too long, the response is still inadequate. Indeed, a very reflective room, such as a reverberant test chamber, would exhibit long decays even in the mid frequency range and although the RT can be quite homogeneous across frequency, such a room would still be considered unfit for sound reproduction. Hence, attempts to correct the modal response must necessarily target modal damping, increasing bandwidth and reducing decay time. This will have a more efficient effect than increasing density.

Finally, if modal density is to be considered as an indication of improved reproduction quality, then the results predicted by the Schroeder Frequency underestimate this, especially for smaller rooms. The use of Fs in such spaces is in itself controversial given that diffuse conditions are never really found in realistic cases [7].

6. REFERENCES

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

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

[3] 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.

[4] 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.

[5] M.R. Schroeder, "The ``Schroeder Frequency'' Revisited," J. Acoust. Soc. Am., vol. 99, May. 1996, pp. 3240-3241.

[6] 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.

[7] F.E. Toole, "Loudspeakers and Rooms for Sound Reproduction—A Scientific Review," J. Audio Eng. Soc, vol. 54, 2006, pp. 451-476.

[8] B.C. Moore, An Introduction to the Psychology of Hearing, Academic Press Inc, 1997.

[9] T. Houtgast and H.J.M. Steeneken, "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria ," J. Acoust. Soc. Am., vol. 77, Mar. 1985, pp. 1069-1077.

[10] L.E. Harris, K.R. Holland, and P.R. Newell, "Subjective assessment of the modulation transfer function as a means for quantifying low-frequency sound quality," Proc. Inst. Acoust. (UK), vol. 28 (8), 2006.

[11] B. Fazenda, K.R. Holland, and P.R. Newell, "Modulation transfer function as a measure of room low frequency performance," Proc. Inst. Acoust. (UK), vol. 28 (8), 2006.

[12] N. Stefanakis, J. Sarris, and G. Cambourakis, "Source Placement for Equalization in Small Enclosures," J. Audio Eng. Soc, vol. 56, May. 2008, p. 357.

[13] D.W. Robinson and R.S. Dadson, "A re-determination of the equal-loudness relations for pure tones," British Journal of Applied Physics, vol. 7, 1956, pp. 166-181.

[14] R.H. Bolt, "Normal Modes of Vibration in Room Acoustics: Angular Distribution Theory," J. Acoust. Soc. Am., vol. 11, Jul. 1939, pp. 74-79.

[15] M.M. Taylor and C.D. Creelman, "PEST: Efficient Estimates on Probability Functions," J. Acoust. Soc. Am., vol. 41, Apr. 1967, pp. 782-787.

[16] M.M. Taylor, S.M. Forbes, and C.D. Creelman, "PEST reduces bias in forced choice psychophysics," J. Acoust. Soc. Am., vol. 74, 1983, p. 1367.