Perceptual thresholds for the effects of room modes as a function of modal decay

Bruno M. Fazenda[a)]
Acoustics Research Centre, University of Salford, Salford, United Kingdom

Matthew Stephenson[b)]
University of Huddersfield, Huddersfield, United Kingdom

Andrew Goldberg
Georg Neumann GmbH, Berlin, Germany

(Received 16 July 2014; revised 12 January 2015; accepted 22 January 2015)

Room modes cause audible artifacts in listening environments. Modal control approaches have emerged in scientific literature over the years and, often, their performance is measured by criteria that may be perceptually unfounded. Previous research has shown modal decay as a key perceptual factor in detecting modal effects. In this work, perceptual thresholds for the effects of modes as a function of modal decay have been measured in the region between 32 and 250 Hz. A test methodology has been developed to include modal interaction and temporal masking from musical events, which are important aspects in recreating an ecologically valid test regime. This method has been deployed in addition to artificial test stimuli traditionally used in psychometric studies, which provide unmasked, absolute thresholds. For artificial stimuli, thresholds decrease monotonically from 0.9 s at 32 Hz to 0.17 s at 200 Hz, with a knee at 63 Hz. For music stimuli, thresholds decrease monotonically from 0.51 s at 63 Hz to 0.12 s at 250 Hz. Perceptual thresholds are shown to be dependent on frequency and to a much lesser extent on level. The results presented here define absolute and practical thresholds, which are useful as perceptually relevant optimization targets for modal control methods.

[http://dx.doi.org/10.1121/1.4908217]

[MV] Pages: 1088–1098

I. INTRODUCTION

Since the 1930s, there has been extensive research regarding the quality of low frequency reproduction in small rooms, and yet, the area remains continually cited as problematic among designers and end users alike. Well documented are numerous techniques which have been applied in an attempt to control this region of the sound-field. Examples include aspect ratio design of the listening rooms,[1–5] specific loudspeaker placement,[6–9] the use of multiple subwoofers,[10,11] complex DSP (Digital Signal Processing) equalization,[12] passive absorption,[13] and more elaborate active methods.[14] These methods each have their own limitations—from reliance on simplistic assumptions, to issues with practicality, cost, and even esthetics. One of the most consistent problems, however, lies with the fact that the criteria by which these methods are evaluated are often *objective* in nature, i.e., the design target is based on an arbitrary value for a metric, for example "the reverberation time should be less than 0.3 s." It is not that the objective criteria are incorrectly evaluated, on the contrary, models and measurements have become increasingly accurate due to greatly increased processing power and better instrumentation. It is that if such an objective measure cannot be shown to be perceptually valid, then even if an optimization scores highly for a given objective metric, an improvement may or may not be perceived (see, for example, Refs. 11 and 15). Worse still, the objective criteria may be too strict—the optimization may in fact "improve" the reproduction quality beyond that which can be perceived. In this case, the additional cost of such a solution would be unjustified.

As will be shown in Sec. II, reducing the modal decay is considered an important optimization in the control of problems introduced by room modes in the low frequency range of listening rooms. Therefore, the aim of this study is to gain a greater understanding of the sensitivity of human hearing to these problems and to define thresholds for the perception of modal effects based on the control of modal decay.

II. BACKGROUND

Low frequency reproduction has become increasingly problematic as modern loudspeakers are able to reproduce lower frequencies, program material has more low frequency content, and smaller listening spaces become the norm due to the increasing cost of available floor area and equipment becoming physically smaller. In such environments, room modes dominate the low-frequency response of the loudspeaker at the listening position.

In terms of a perceptual response, early investigations focused on the audibility of resonances. Bucklein[16] showed that upward deviations in the magnitude response are more

[a)]Author to whom correspondence should be addressed. Electronic mail: b.m.fazenda@salford.ac.uk

[b)]Prior to 2012, Matthew Stephenson was known as Matthew Wankling, as cited in a number of references here.

audible than downward ones, however, with the exception of two resonances at 85 and 150 Hz, the work focused on resonances above 200 Hz. Further study by Toole and Olive[17] also focused on frequencies higher than 200 Hz, albeit in greater detail, investigating the threshold of audibility of resonances as a function of frequency, Q-factor, relative amplitude, onset time delay, program material, listener hearing performance, loudspeaker directivity, and reverberation added during recording or reproduction. The study demonstrated that temporal changes and the reverberation time at higher frequencies affect the threshold of detection of resonances.

Studies into the detection of resonances at lower frequency were carried out by Olive *et al.*[18] at frequencies between 63 and 500 Hz. The main results indicated that, using pink noise as the test signal, the detection thresholds decrease with increasing Q-factor. It was also shown that for broadband steady state signals, detection worsens as frequency decreases with exception of lower Q resonance detection, which appears to be independent of frequency. An interesting result revealed that temporal aspects of the signal are important in the detection of resonances and that, when transient signals (pulses) are used, the detection thresholds actually decrease considerably at higher Q values ($Q = 30$). Additionally, and under such conditions (transient signal, high Q), antiresonances are as detectable as their equivalent resonances. These results suggest that, in the presence of music signals, which are in their nature composed of many transients, room resonances may impart a much different perception when compared to broadband steady state signal excitation. Further research has corroborated this idea that temporal decays may be more significant in determining perceived low frequency quality than the more traditionally used modal distribution metrics.[19]

Efforts to optimize low frequency reproduction in rooms attempt to achieve an homogeneous power spectral response (i.e., flat across frequency) and a reduction of the resonant behavior of the modes, thus reducing their decay time. Whether a simple passive absorption approach or a more complex active room equalization is attempted,[20,21] the complexity and cost of such methods may be greatly reduced by using perceptually valid decay thresholds as guiding targets for the final response.

Sophisticated modal control methods seek to invert the complex frequency response to achieve a close approximation to a unity transfer function (no change to magnitude or phase) within a certain bandwidth of interest.[22–26] It is reported that these techniques can lead to side effects in the temporal response, which may be perceived as degradations of the response depending on the severity of the equalization attempted.

In active modal control implementations, decay time is identified and an equalization technique is implemented that attempts to reduce the pole radii of the modes in the overall transfer function.[20,21] An alternative implementation finds peaks in the low frequency response, assumes they are due to resonances, and introduces parametric equalization filters which flatten the spectral response thereby making resonances less audible.[27] However, this method can actually increase the decay time making the problem more audible as one is adding a filter which has its own resonant behavior, or audible beating is heard if the resonance and filter frequencies are slightly different.

Most of the research mentioned hitherto typically relies on physical metrics, measured directly or derived from the room response, to gauge the overall success of modal control methods. In more recent studies, the notion is emerging that perceptual thresholds for modal decay can help reduce the complexity of the modal control system, facilitating a faster optimization, with a simpler and therefore more practical implementation. Karjalainen *et al.*[28] studied the perception of decay time at a number of frequencies. Single resonances, representing room modes, were added to the driving signal of a single loudspeaker in a room with its own modal soundfield and mid-frequency reverberation. It was found that at typical listening levels the threshold for modal decay time increased from about 0.3 s at 200 Hz to 0.4 s at 100 Hz. However, when testing at 50 Hz, subjects observed no noticeable differences for decay times of up to 2 s, which opened up the question of whether it is worth attempting to perform modal correction at these very low frequencies. The fact that a single resonance was being controlled and that no natural interaction or variation of modal effects in the room was implemented suggests the effects of these on modal decay thresholds were not tested. Although the subjective response to individual resonances is of interest to establish a perceptual basis of detection, within the context of a room, many resonances exist, often interfering with each other, producing a complex time-frequency response. Closely spaced resonances have been shown to cause additional effects such as beating in the overall decay pattern and a commensurate reduction of the audibility threshold.[29] The effects of a full room response were considered by Avis *et al.*[30] who attempted to define thresholds of modal Q, by modeling a listening room with "bi-quad" filters at frequencies down to 34 Hz. The Q factor of these filters could be dynamically varied and used as the independent test variable. The "filter" room models were convolved with real music signals and the resulting audition samples presented over headphones. Listening tests showed an absolute threshold of $Q = 16$ below which resonances were inaudible. Modal decay time is related to Q through the simple formula in Eq. (1), and so, this result can be extrapolated to indicate decay time thresholds of 1.1, 0.5, and 0.2 s at 32, 63, and 125 Hz, respectively. However, the independent variable on those tests was the Q factor which, when applied equally to all modes in a given test sample, results in monotonically decreasing decays across the frequency range under study. Consequently, the extrapolation of obtained Q factor thresholds onto frequency dependent decay thresholds cannot be assumed since we do not know on which frequency the subjects were basing their decisions.

$$T_{\mathrm{modal}} = \frac{2.2Q}{f}. \tag{1}$$

In a recent subjective testing of modal control systems, the Controlled Acoustic Bass System setup,[14] while revealed as a "good" control system, was not perceived as significantly better than more simple solutions which reduce the decay time to a lesser extent, but would be "cheaper" to implement.[11] The underlying message stemming from such a result is that the reduction of modal decays beyond a given threshold might be unnecessary for a perceptual improvement of reproduction quality, which is the same conclusion drawn by Karjalainen *et al*.[28]

It follows then that there is a need for further examination of the thresholds of decay at low frequencies which consider realistic listening scenarios—rooms with multiple interacting resonances—with representative program material. Of particular interest in the study presented here is how comparable are the results for absolute thresholds using controlled test tones, and more "natural" thresholds obtained in the presence of real room acoustic conditions and musical signals. Consequently, this study has been designed to include detection tests for each of these cases, with a comparison between the two sets of thresholds forming part of the discussion.

III. DEFINING TEST SIGNALS AND MODELING THE ROOM DECAY

"Modal decay," the dependent variable under test, is defined here as the time taken for an individual resonance to reduce in amplitude by 60 dB after excitation is removed. It is important to note that while this concept is borrowed from the definition of reverberation time, it cannot strictly be called "reverberation time" as it is not a diffuse field parameter.

The perceptual *threshold* is defined such that if a mode has a decay time shorter than the particular threshold, the individual effect of that mode on overall audio quality would typically be inaudible to the majority of listeners.

This paper documents two threshold tests: (1) the first test deals with absolute perception of decay time for single resonances, termed here *artificial* stimuli and could be considered as a direct measurement of the perceptual thresholds of modal decay since, as a single frequency test, they do not include any other side effects; and (2) those thresholds within the context of real rooms and music signals, henceforth called *natural* stimuli, which provides a more ecologically valid test since perceptual effects of both tonal and temporal nature will be introduced. The complexity of the latter tests requires a slightly modified methodology, and this is discussed in due course.

A. Test signals—Artificial

In order to determine perceptual thresholds based on modal decay, suitable test signals must be generated. A number of signals are commonly used in psychoacoustic testing, such as white and pink noise, pulses, logarithmic and linear sweeps, and pure tones. These signals can be considered representative of differing components of a natural signal. For example, pulses were used by Olive *et al*.,[18] who suggested that they are helpful in revealing the audibility of decays when a musical signal contains transient sounds. Similarly, single tones are representative of harmonic elements, while noise can be used as a controlled artificial stimulus to reveal perceptual responses to non-melodic musical elements.

The construction of an ideal sine burst is discussed in detail by Goldberg,[31] but the method is also briefly presented here for completeness and clarity. To find modal decay thresholds, Goldberg originally intended to use pure tone sine bursts where the end of a simple sine tone would have been amplitude modulated with the exponential decay curve of a single frequency room mode under test. However, it was discovered that spectral spreading caused by the switch-off transient at the end of the sine burst meant listeners heard a side effect (bumping sound) that was not related to the modal decay, thereby leading to very low and incorrect threshold values. A half-cosine window (Hann) was applied to the end of the sine burst as a fade-out window and tested to be as short as possible to avoid the audible spectral spreading side effect and not affect the modal decay. The fade-out window length is 30 ms window for frequencies above 100 Hz, whereas three cycles of the sine tone are needed below 100 Hz, for example, 47.6 ms at 63 Hz. These fade-out window times are well below the decay times observed within rooms and also much shorter than the thresholds suggested by previous studies. Starting at the end of the sine burst where the fade-out window starts, a modal decay is then overlaid onto this perfect sine burst to simulate the modal decay being tested. This allows the modal decay to drop all the way to 0 s without any other side effects becoming audible. A fixed 50 ms Hann window is also applied to the beginning of each burst to avoid switch on transients and careful attention is also paid to the audio reproduction equipment to avoid any other spurious noises.

The decaying sine waves simulate a resonant system directly and, therefore, each test can consist of a single burst at the required frequency. For the determination of these thresholds, other signals in the test stimuli are not required because they introduce additional effects such as simultaneous and post-masking.[32] Additionally, extra stimuli would increase the demand on subjects and, as the goal is to obtain baseline absolute thresholds of modal decay before continuing to determine thresholds in more realistic scenarios, other stimuli which are likely to represent elements within the music were considered unnecessary.

This describes the artificial test signal used in this paper and, together with the listening test method described in Sec. IV, leads to a direct measure of the perceptual threshold of the modal decays.

B. Test signals—Natural

In the case of artificial stimuli, single decaying pure tones provide the ability to directly determine the effect of frequency on the perceptual thresholds. However, achieving this frequency dependency becomes more complex when considering more ecologically valid contexts with room responses and music stimuli. A methodology based on the auralization of music samples through a room model was developed for this.

Two music samples were used for the determination of decay thresholds. The music samples will henceforth be

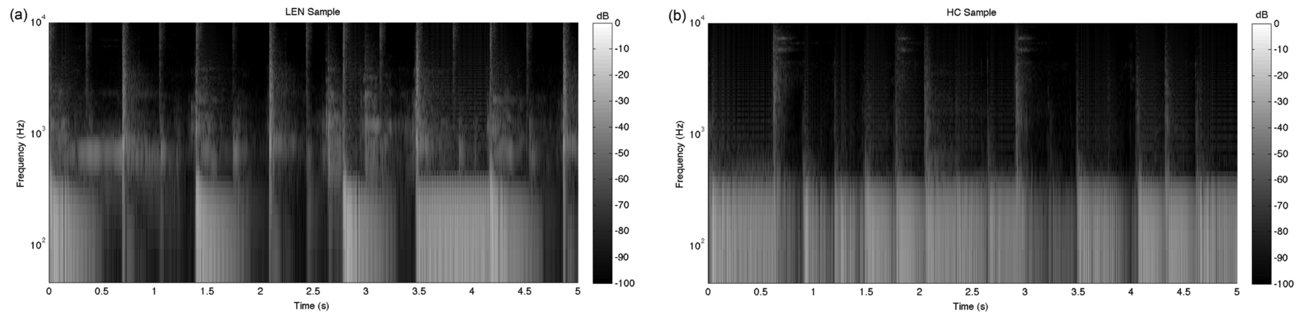

FIG. 1. Spectrograms of the two music samples.

referred to as LEN and HC. They represent music with both short, well defined low frequency content, and also resonant acoustic bass notes with a naturally longer decay envelope:

(1) "LEN" may be considered to be in the "pop" genre.[33] It is around 6 s in length and contains a number of sparse bass notes with short attacks and decays, and a sparse kick drum rhythm. Between the bass notes (or drum hits) there are long gaps wherein the modal decays can be clearly audible. This sample has been used with success in previous research.[30]
(2) "HC" is of similar length and may be considered to be in the "jazz" genre.[34] It is a solo double bass refrain, with a greater range of pitches than LEN. It is therefore likely to excite a wider band of frequencies in the experimental region. It is observed that due to the natural resonant behavior of the double bass, some notes in the sample have longer decay times than some of the modal decays under test.

Figure 1 shows the spectrograms of the two music samples.

C. Room model auralization

The auralization process is based on a low frequency room model, allowing modal decay time of the modeled responses to be quickly and accurately modified. Figure 2 shows a schematic for the auralization process. Essentially, the auralization is broken into two frequency ranges:

(1) The higher frequencies come from the original music signal high pass filtered at a chosen crossover frequency (see Sec. IV) using a fourth-order Butterworth filter. It therefore contains no specific "room" content. No high frequency room model was included in the auralization as this would distract from the main purpose of the experiment.
(2) The low frequency room response was modeled using the Green's function for a cuboid room, described in detail below, and then convolved with a down-sampled version of the input audio (fs = 2000 Hz). Headphone equalization for a pair of Sennheiser HD650 headphones consisted of a 3000 tap finite impulse response filter corresponding to the inverse of the low frequency transfer function of the left ear piece measured on a HATS. The target response was a flat magnitude response up to 2 kHz.[35] An overall energy level calibration between the original low frequency region of the audio sample and the new modeled version was applied to maintain the original artistic balance of the production. The result was then up-sampled back to 48 kHz. Finally, this equalized and modeled low frequency part of the original signal was low pass filtered, using a fourth-order Butterworth filter, ensuring that the low and high frequency regions of the audio sample cross over with a flat magnitude response.

The high- and low-frequency regions are then summed to produce the output signal which is presented over headphones.

In order to measure frequency dependent thresholds, the auralizations were presented to subjects with a variable "cut-off" frequency applied to the modeled low frequency room responses. The variable frequency corresponds to the

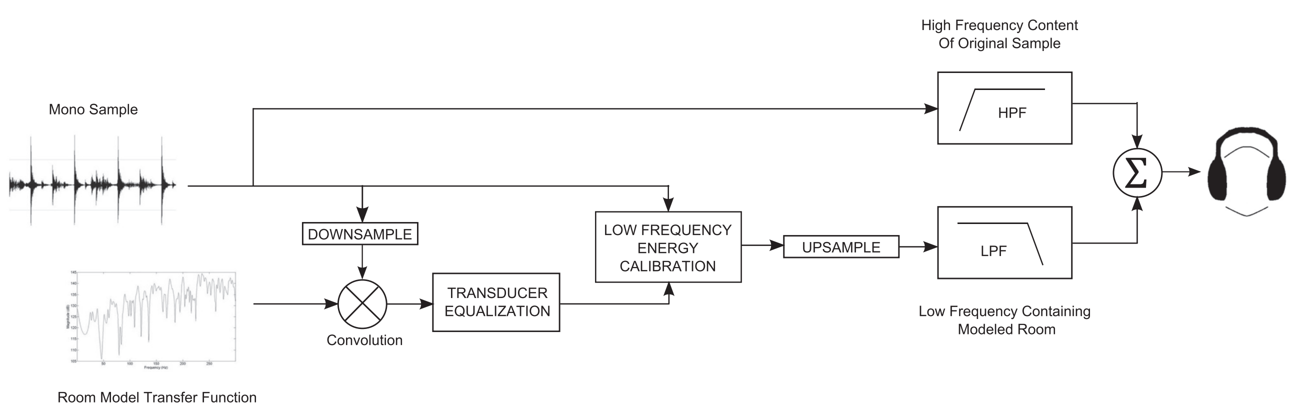

FIG. 2. A schematic of the auralization process.

crossover between (a) the room model convolved with the low frequency region of the music sample and (b) the original music sample (see Fig. 2). Three crossover frequencies were tested—63, 125, and 250 Hz. This allows a "cumulative frequency dependency" to be observed. For example, where the cut-off is 63 Hz, the threshold revealed is applicable up to this frequency. This will mean that with a 125 Hz cut-off frequency, the samples will also include the modal decays below 63 Hz. It has been noted that modal decay thresholds decrease with increasing frequency[18,28,30] as does the sensitivity of hearing.[36] Thus, for a model with constant decays across the frequency range tested, it is posited here that the thresholds obtained will refer to the highest modal frequency region modeled, thus superseding any thresholds applicable for lower frequency cut-off points.

A generic cuboid shaped room with low damping can be adequately modeled by its Green Function, also known as a modal decomposition model, described by Kuttruff,[37]

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

where c is the speed of sound, Q the source strength, ρ the density of air, and w_n values are angular modal frequencies, which are defined (in rads^{-1}) from the following:

$$\omega_{xyz} = c\pi\sqrt{\left(\frac{x}{L}\right)^2 + \left(\frac{y}{W}\right)^2 + \left(\frac{z}{H}\right)^2}, \qquad (3)$$

where x, y, and z are the number of half wavelengths between surfaces and L, W, and H the distance between surfaces. Cartesian coordinates are used throughout.

The modal decomposition model was implemented to generate room impulse responses where the decay times of the modes can be controlled directly. For this experiment, all modes in the model are set to decay at the same rate, which means that their Q factor increases with decreasing frequency [Eq. (1)]. This modeling technique has been shown to be successful in previous subjective studies.[19,38] The other input parameters were kept constant: room volume, 100 m^3; dimensions, $x = 6.97$ m, $y = 5.32$ m, $z = 2.69$ m; source position, front-left-bottom tri-corner (modeled as a point source); receiver position, $x = 3.16$ m, $y = 1.97$ m, $z = 1.3$ m; and frequency resolution, 0.12 Hz. These values were chosen to represent a typical, well designed, listening environment adhering to a good room ratio. In other words, a room without any major acoustical issues or atypical metrics when compared to listening environments typically found in professional audio and research facilities. For all cases modeled, the summation includes modal frequencies up to 300 Hz to ensure the residues of modes above 250 Hz are adequately taken into account in the response.

All modes up to the cut-off frequency are modeled with identical decays. This differs from the experimental work of Avis *et al.*[30] in their investigation of the threshold of modal Q, which was kept constant for all modes within a given test sample, resulting in differing decay times across frequency. In the test presented here the decay time was controlled through the analytical model's damping parameter δ in Eq. (2). The required alpha (α) for a given decay time was obtained through the use of Sabine's equation relating reverberation time (T_{60}) to the absorption coefficient, α (see Morse[39]). It is therefore possible for the modeled decay to be dependent on both frequency and absorption at each boundary, although this was simplified in this model by attributing a single α for all surfaces. The impulse responses produced from an inverse Fourier transform of the resultant complex pressure vector may be verified for decay time characteristics using a Schroeder backward integration plot. Figure 3 shows the model output in the form of a cumulative spectral decay and integration plots for each of the three cut-off frequencies tested, at a modeled decay of 0.5 s.

The modal decomposition model assumes infinitely rigid boundaries which becomes invalid when damping is high.[39] In such cases, the model accuracy decreases and deviations occur in the calculation of the angular frequencies. This is an often cited problem of using such simple analytical models of rooms when high damping cases are being investigated, as is the case here. However, in this study, the model is being used to adjust the rate of modal decay rather than to generate a precise replication of the response for a

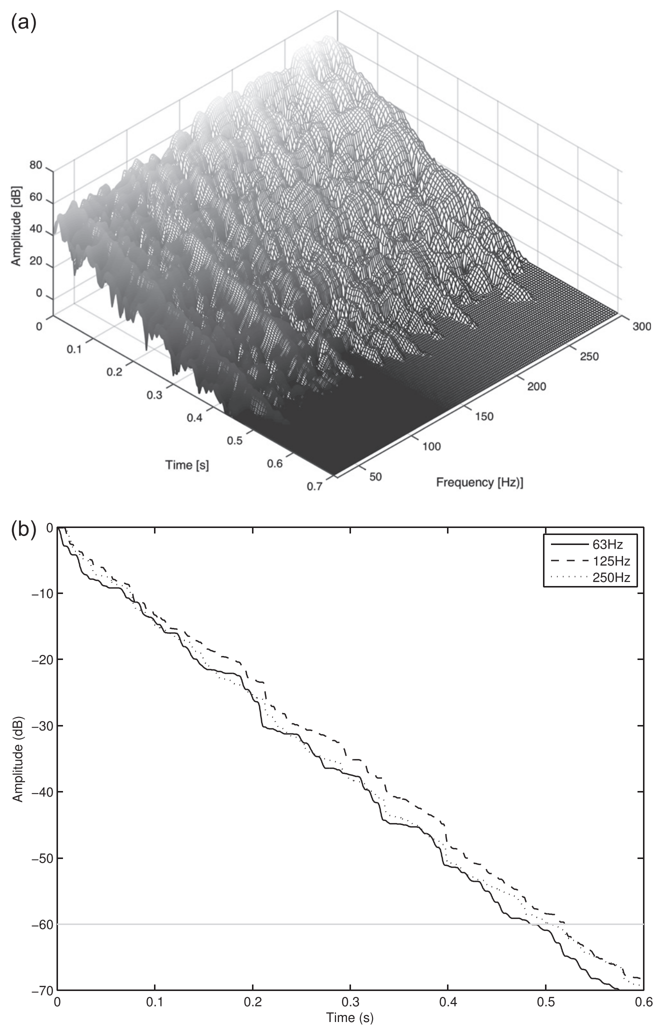

FIG. 3. Model output at 0.5 s modal decay, (a) cumulative spectral decay and (b) Schroeder integration plot at each tested frequency.

specific room. Under such premise, it is argued that the model does continue to provide a *general case* of the room response and is therefore adequate for the generation of the audition samples.[40] Also missing is a noise floor which could increase threshold values in real rooms, but since our goal is to measure absolute thresholds for modal decay, the addition of a noise floor would become an unhelpful confounding factor in the experiments and in high quality recording studios the noise floor is very low anyway.

IV. MEASURING THE THRESHOLD

With the two test signals and auralization model defined, we now consider the methodology for determining the perceptual thresholds.

A. Test variables

In order to define absolute thresholds, two sine bursts were produced (1) the reference, with no decay (but short Hann window to prevent spectral leakage as stated in Sec. III A) and (2) the variable signal with decay time determined by the testing method.

Five sine burst frequencies were tested—32, 63, 100, 150, 200 Hz—at two reproduction levels of 70 and 85 dB sound pressure level (SPL) (linear, unweighted, at the eardrum). These levels were chosen to be "quiet" while remaining audible, and "loud" but without introducing acoustical distortion from the headphones.

Testing of natural stimuli was set at three cut-off frequencies of 63, 125, and 250 Hz, but because of the nature of the signals, reproduction levels of 70 dB SPL were considered too quiet, and so levels of 75 and 85 dB SPL (linear, unweighted, at the ear drum were used) (see Sec. IV E).

The dependent variable is the decay time, and we are trying to find the decay time where no difference is noticeable between the reference signal, with no decay, and the signal which includes a decay.

B. Method

The Parameter Estimation through Sequential Testing (PEST) was first defined by Taylor and Creelman[41] as a decision based method to quickly and efficiently converge on a particular threshold. The method is based upon a routine with a set of rules, run after each subject decision (a *trial*), dependent upon the subject's previous responses. A successful *run* is composed of a number of consecutive trials converging onto the desired threshold measurement. Both Avis *et al.*[30] and Goldberg[31,32] employed this method in threshold testing.

In the work presented here, we are interested in defining the point where a difference between a reference sample and a variable sample are no longer detected, meaning the effects of modes are no longer perceptible. The PEST method has been modified with a criterion-free ABX test at each trial to determine whether the subject can, without doubt, hear a difference, thus determining a positive or negative response for the convergence routine to calculate the next decay rate value. Subjects are therefore asked to determine which sample, A or B (randomized between reference and variable) is X, by means of a graphical user interface. In order to verify that a listener can correctly judge sample X, they must answer correctly three times. If three consecutive identifications are made, the routine is fed the positive result that the subject could indeed hear a difference between the samples, while just one incorrect answer signals a failure to do so; and the routine is updated accordingly. The requirement of three consecutive correct answers reduces the probability of the subject guessing to 12.5%. Following the outcome of the trial, the determination of the next decay time is carried out using the step size rules suggested in Taylor and Creelman.[41]

The maximum number of trials per run was set to 30 with a maximum allowed number of six reversals. The termination of a run and subsequent estimation of the threshold is obtained when a new step size is required which falls below a pre-defined minimum. The minimum step size for our tests was defined, after a number of pilot tests, as 0.025 s for artificial stimuli and 0.05 s for natural stimuli. A maximum decay time was set as 2 s and, if the decay time should ever drop below 0 s as a result of the convergence rules, it was reset to 0.025 s. With hindsight, the minimum defined step size corresponds to a maximum error of around 12.5% to 25% of the lowest measured thresholds for those subjects reaching the minimum step size in their trial runs. As the number of reversals was kept low to avoid fatigue, it is possible that a listener's run was terminated before reaching the minimum step size, effectively adding noise to our measured data across the panel. However, the use of 10 and 16 listeners in the test panels reduces this noise by a factor of 9 to 12 dB, thus restoring confidence in the measured data. This is reflected in the reasonably tight confidence limits obtained for the artificial test signals presented in the Sec. V. The wider confidence limits for the natural signals reflects the harder task asked of the subjects rather than inaccuracy in the test method.

The frequency and replay level of the first and last tests were fixed for both tests (100 Hz/85 dB for artificial and 125 Hz/85 dB using the LEN sample as the natural stimuli). These may then be compared in order to study the negative effects of subject fatigue, the positive effects of learning and general listener reliability. All other tests were randomized between subjects so as to avoid presentation bias.

C. "First impression" PEST/ABX

Tests using natural stimuli were conducted several months after the artificial stimuli test and included only four common subjects. This avoided any bias between the two tests. A pilot test was run using the same PEST/ABX methodology as detailed above in an attempt to reveal corresponding perceptual thresholds with music. However, it became apparent that, when testing using music stimuli, the routine failed to converge. Subjects were detecting differences between the test and reference samples right down to decay values very close to the reference. The routine then caused a drop *below* the reference and subjects continued to report differences forcing the sample's decay to be reduced further, thus never converging to a threshold. It was however observed that perceiving the differences within the context of the ABX test was not simple and could only be performed if subjects were allowed to repeatedly and instantaneously compare between the samples. Additionally it was noted that, after a number of comparisons, subjects would invariably find one feature within the samples which allowed them to detect a difference unless the two samples were identical. This clearly was not revealing a useful threshold that corresponded to room conditions. Indeed, these effects had already been observed and discussed in Fazenda and Wankling.[29] Furthermore, this type of instantaneous AB comparison is not typically found in real world scenarios, and thus, the ecological validity of the test with natural stimuli would become questionable.

In light of these observations, a further modification to the PEST/ABX method was made to reveal a "first impression" threshold. The routine rules remained identical to the artificial stimuli test. However, the ABX section allowed only a single audition of each of the A, B, and X samples. Once the sample had been auditioned, the corresponding play button was disabled and an answer required. This still allows subjects to use short term memory for comparison as the samples can all be played within an interval of 15 s. In addition to being able to indicate whether sample X was either A or B, an additional "Unsure" button was made available on the interface, which had the same effect as answering incorrectly—that is, not being able to reliably determine sample X. This offers a "get out clause" that removes the frustration from the subject of answering incorrectly.

D. Listening conditions, equipment, and listeners

A portable set-up enabled testing to be carried out in two locations, a quiet listening room conforming to the standards for small impairment listening tests (BS-1116), and a semi-anechoic room. Both rooms have a background noise sufficiently low that prevents this from being a negative or biasing factor in the results.

A dual monitor set-up was used which allowed the test administrator to view the test progress and enforce breaks as necessary at convenient intervals not more that 15 min apart.

All listeners were members of staff or students from the University of Salford's Audio and Acoustics department. A total of seventeen subjects participated in the artificial stimuli tests, and ten in the natural stimuli tests (of which four subjects were common to both tests). They were specifically instructed that the test was searching for a decay threshold, and it was a change in this parameter which should be listened for. While fewer subjects participated in the natural stimuli tests, each was specifically invited and had participated in a number of similar tests assessing the same program material in a variety of situations over both loudspeakers and headphones. The group of subjects who took part can therefore be considered as a panel of experts in this topic.

E. Level calibration

For artificial stimuli, consisting of five discrete frequency sine bursts, it was possible to calibrate the replay levels at each individual frequency. This was achieved by calibrating the headphone output with a B&K HATS system and a Norsonic Sound Analyzer for replay levels of 70 and 85 dB SPL.

Although audible, the sine burst at 32 Hz and 70 dB replay level was perceived as very quiet. It was difficult to perceive differences in decays, and as a result, this lowest frequency was only tested at the 85 dB replay level in order to reduce strain on the listener.

Using a sound level meter it is possible to determine the sound pressure level of the artificial tones. However, when using musical signals, it is more challenging to report a "loudness level." The auralized music samples were therefore calibrated by measuring the loudness level as defined in ITU-R BS.1770-1.[42] This standard refers to an "audio program loudness" which has units which are equivalent to a decibel level. In order to calibrate, a standard 1 kHz sine tone was used as a reference. The tone was played through the headphones onto the B&K HATS. The level was measured with a sound analyzer and the output from the soundcard adjusted until the level was either 85 or 75 dB, respectively. This tone was then passed through the ITU loudness algorithm to obtain a reference loudness value for the 1 kHz tone appearing at the ears at 75/85 dB. Any auralized music sample can then also be passed through the algorithm and adjusted by the appropriate gain factor such that the overall sample loudness can be considered perceptually similar to the 1 kHz tone. Samples were calibrated using linear weighting.

V. RESULTS

A. Artificial stimuli thresholds

Subject reliability was measured by considering the number of trials which failed to converge and the consistency of thresholds between each subject's first and last tests as defined in Sec. IV B. One subject's results were removed from subsequent analysis because there was a large discrepancy between their first and last test results and two of their test runs failed to converge.

The mean time taken for the tests was 29 min (including breaks) with an average of 3.6 min/convergence. The average number of trials needed for convergence on a threshold was ten.

Figure 4 shows the mean detection thresholds and 95% confidence interval (CI) at each frequency and replay level. The thresholds appears to decrease consistently with frequency until about 100 Hz, where they converge to around 0.2 s. At 63 Hz, thresholds are higher at low replay levels, which is an expected perceptual behavior arising from the decreased sensitivity to level as frequency decreases.[36]

To explore the data set further and determine the statistical significance of these thresholds, a two way repeated measures analysis of variance (ANOVA) was performed with Replay Level and Frequency as the factors under analysis. The data collected at 32 Hz was not included in this analysis due to the missing data at the 70 dB audition level. In this data the independent variables explain 55% of the variance in the data.

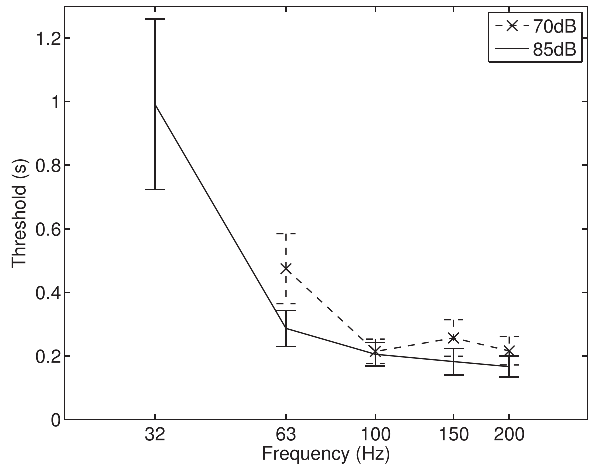

FIG. 4. Mean decay time thresholds and 95% CI for sine bursts.

A Mauchly's test of Sphericity was applied and since upon this data set sphericity could not be assumed for the factor *Frequency*, or the interaction between the two factors, a Greenhouse-Geisser correction has been applied to obtain the significance level p.

The ANOVA shows significant interaction effects between the two factors and also highly significant main effects for each factor (Replay Level × Frequency: *Greenhouse-Geisser correction*; $F(1.6, 23.99) = 6.038$, $p < 0.05$, $\eta^2 = 0.07$, Replay Level: $F(1, 15) = 28.26$, $p < 0.01$, $\eta^2 = 0.11$; Frequency: *Greenhouse-Geisser correction*; $F(1.87, 28.02) = 27.64$, $p < 0.01$, $\eta^2 = 0.38$). The effect size for the interaction between replay level and frequency is small ($\eta^2 = 0.07$) and explains only about 7% of the variance. Such weak interactions are difficult to interpret, and in this case, it appears to be associated with the steeper decrease in threshold values from 63 to 100 Hz for the 70 dB replay level, when compared to the 85 dB replay level. The effect size for Replay Level is also small at $\eta^2 = 0.11$. The largest effect size observed is for Frequency at $\eta^2 = 0.38$. In order to identify where the significant differences between the thresholds at each frequency lie, a *post hoc*, Bonferoni corrected, multiple comparison was carried out and a statistical significance was obtained for each of the compared pairs. The *post hoc* tests show a highly significant difference ($p < 0.01$) between 63 Hz and all the other tested frequencies and no significant differences ($p > 0.05$) between any of the other pairs of frequencies. This strengthens the argument that perceptual thresholds of decay time at frequencies of 100 Hz and above remain consistent while, below this, thresholds increase. This result agrees with the study by Karjalainen *et al.*[28] and is further discussed in Sec. VI.

B. Natural stimuli thresholds

As with the artificial stimuli analysis, each subject's performance was evaluated. In this case, no subject data was removed, and the mean time taken was 32 min, with a convergence time of 2.7 min.

Figure 5 shows the mean and 95% CI for the two samples at each replay level. Thresholds measured follow a similar trend to those obtained for the artificial stimuli with thresholds decreasing with increasing frequency. The variance of the data has increased considerably, demonstrating most probably the added difficulty in detecting decaying modal energy in the presence of masking caused by subsequent musical events.

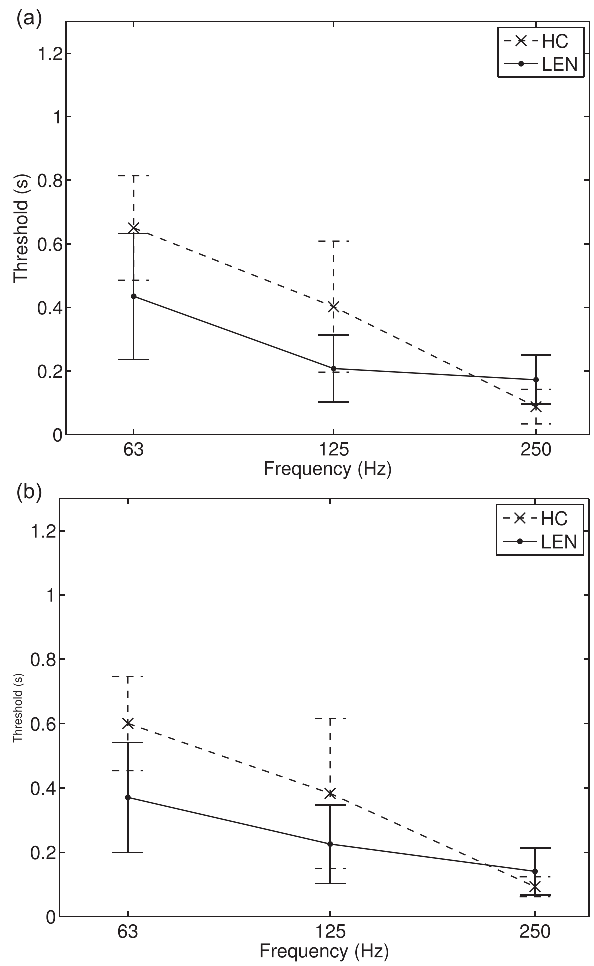

FIG. 5. Mean decay time thresholds and 95% CI for two music samples, (a) 75 dB and (b) 85 dB replay levels.

A n-way repeated measures ANOVA was carried out including the three factors, Replay Level, Frequency, and Sample. In this data, 69% of the variance is explained by the independent variables and their interactions. The ANOVA results reveal a significant but weak first order interaction between frequency and sample (Frequency × Sample: $F(2, 12) = 11.873$, $p < 0.01$, $\eta^2 = 0.09$). This is explained by the steeper increase of thresholds with decreasing frequency for the HC sample, when compared to the LEN sample. The former sample has been reported as more difficult when assessing audibility of modal decays. Significant main effects are reported for Frequency, with a large effect size, accounting for 51% of the variance observed [$F(2, 12) = 113.76$, $p < 0.01$, $\eta^2 = 0.51$], and Sample, with a weak effect size accounting for 7% of the variance [$F(1, 6) = 6.77$, $p < 0.05$, $\eta^2 = 0.07$]. Replay Level was not significant [$F(1, 6) = 0.31$, $p > 0.05$, $\eta^2 = 0.002$]. The same trend obtained for artificial stimuli is observed, i.e., thresholds decrease with increasing frequency.

There were only two levels for the factors Replay Level and Sample, thus no *post hoc* tests are required. However, the ANOVA shows significant differences for the factor *Frequency* and, therefore, a Bonferroni corrected, *post hoc* multiple comparison was performed in order to determine where the difference lies. Highly significant ($p < 0.01$) differences between thresholds at all pairs of frequencies were found.

VI. DISCUSSION

The implications of the findings from the two listening tests are now discussed.

A. Comparison with previous studies

It is interesting to compare these results with those of previous tests. First, there are the initial results from Goldberg's small scale threshold testing,[43] which was primarily conducted to confirm the success of the PEST method for obtaining such decays. At 32 Hz and 85 dB reproduction, thresholds reported in Ref. 43 were around 1.35 s, dropping to around 0.5 s at 63 Hz, with values of around 0.1 s at 80, 125, and 200 Hz. The results presented here show lower threshold values, however, it was observed in Goldberg's experiment that the test signal introduced some masking suggesting the true threshold values were likely to be lower. Karjalainen *et al.*[28] have also reported low frequency decay thresholds. They report a fairly constant 0.2–0.3 s decay time threshold down to 100 Hz, at which point the threshold increases rapidly. Comparison with the thresholds presented in Fig. 5, particularly for the LEN music sample, shows good correlation with Karjalainen's study.

Finally, assuming the thresholds of Q reported by Avis *et al.*[30] can reliably be used to extrapolate approximate decay thresholds, values of 1.10 s at 32 Hz, 0.56 s at 63 Hz, and 0.28 s at 125 Hz were found. The results of Avis *et al.* are slightly higher than the thresholds observed in this paper when using artificial stimuli (Fig. 4), particularly at 32 and 63 Hz, but well within the variance range when compared to thresholds established here using musical stimuli (Fig. 5). Considering that Avis *et al.* have used music stimuli for their studies, the latter comparison is correct.

B. Influence of frequency, level, and stimuli on perceptual thresholds

This study supports earlier findings about detection of modal decays being frequency dependent. A knee point at around 100 Hz was found when testing with artificial stimuli, which is not evident when testing with music, although the fewer data points collected along the frequency continuum for the latter tests might play a part in hiding the true shape of the threshold line.

Results reported here also show that artificial stimuli are dependent on audition level. A significant interaction between level and frequency reveals that thresholds measured at the lower replay levels have a greater increase toward the lower frequencies, which is more obvious below 100 Hz. Above this frequency the effect of replay level does not appear to be significant. It is likely that the rise of thresholds at lower levels is related to the natural drop in hearing sensitivity with decreasing frequency combined with post-simultaneous masking which will "hide" modal decays after the exciting sound has just ceased.[44,45] A common cited masking threshold graph can be seen in Fig. 6. Three temporal decays with times of 100, 150, and 200 ms have been added as dotted lines, and one can see that, for 150 + ms, post-masking has little effect. It should be noted that the post-masking threshold curve shown on this graph was measured using a probe tone (a short sine burst) with a delay sometime after the masker had stopped. Additionally, these tones are not the same as a continuous temporal decay and so it may not be appropriate to use this masking data in this way. These probe tones are usually higher (kHz) not lower (<200 Hz) frequencies. This also assumes that post-masking is the same at all frequencies, which is unlikely, and one should also note that masking is dependent on the duration of the masker. We saw higher thresholds for lower frequencies and also a level dependency. This could be due to longer post-masking or some other hearing effect, but we do not seek to explain the physiological reasons any further here and leave that to a study seeking to develop an auditory model.

The combination of these aspects (masking and absolute threshold of hearing) reduces the time over which modal decays can be audible and so we see a minimum threshold of 0.2 s with an increase toward lower frequencies. The thresholds obtained at a higher replay level of 85 dB are thus more useful in defining absolute thresholds for critical listening scenarios, such as in recording studios where individual instruments are soloed and their sounds manipulated. The lower confidence limits on this data (2.5%) can be further used as an indication of the levels at which the most acute listeners are able to detect modal effects.

Interestingly, results for music stimuli show a weak but significant interaction found between frequency and sample. This confirms previous understanding that the characteristics of music, its tonal, transient and temporal

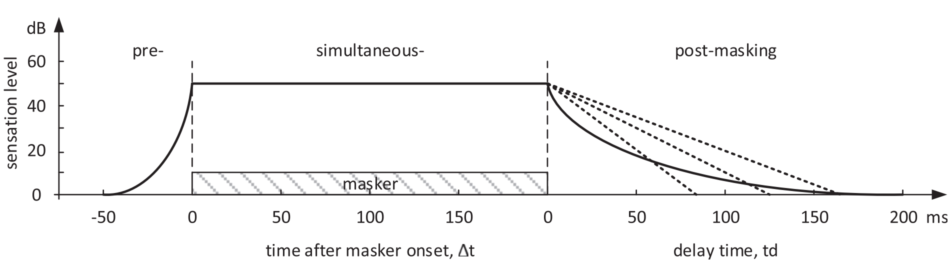

FIG. 6. Pre-, simultaneous- and post-masking curves with added temporal decays represented as dotted lines [adapted from Fastl and Zwicker (Ref. 45)].

content, play a part in revealing modal problems in rooms.

C. Mapping modal thresholds

The results of this paper indicate two levels of perceptual effects of modes:

(1) Testing with artificial stimuli, where one single frequency is auditioned, reveals absolute perceptual thresholds of decay which are based solely on the decay of the modes and how audible these are with respect to post-masking and the threshold of hearing.
(2) Testing in realistic scenarios, with complex modal sound-fields and music stimuli, the thresholds are not solely related to the temporal decay of the modes, but they can also be perceived as tonal variations, or coloration. When instantaneous comparison between two conditions is allowed, discrimination is possible based on these small colorations even though the differences in the sound are solely attributed to differences in the decay of the modes. On the other hand, under realistic conditions typically found in audio reproduction environments, where instantaneous and repeatable comparisons are not possible, the detection of modal effects is more difficult and reverts to a construct of both tonal *and* temporal effects, where variations of the latter are clearly responsible for eliciting different percepts and for which a perceptual threshold exists.

On the basis of the results obtained here, and in the context of realistic reproduction scenarios, it is possible to draw a map defining perceptual regions for modal effects as a function of frequency and modal decay (Fig. 7). Figure 7 shows four regions identified by this study. A lower boundary defined by thresholds obtained with artificial stimuli reproduced at higher listening levels. Below this threshold, it is unlikely that modal problems will be detected and any control methods employed to correct them are likely to be perceptually meaningless. Above this region lies a range of acceptable decays where the reproduction of music signals is not significantly impaired by modal problems. Modal control in this region is likely to be perceived as a decrease in modal activity but revealed only in the presence of carefully selected stimuli and under instantaneous comparison between the before and after conditions. At modal decays well above the thresholds measured with music stimuli (averaged over the two levels tested) it is highly likely that their perceptual effects will be obvious. Under these conditions, attempts to control modal activity will lead to a perceptual improvement. It would then be up to the room designer/user to decide how to bias the decision between having sufficient decay correction for an expensive perceptually "perfect" reproduction on one hand, or a more economic but acceptable amount of audible coloration on the other hand, i.e., a classic cost-benefit analysis that will change from one application to another. Finally, for those with the space, budget, keenest hearing (self-proclaimed "golden ears") and for the most critical listening tasks, the shortest audible modal decay has been derived from the lower 2.5% confidence interval of the measured data and is shown as the top of the darkest shaded region in the graph.

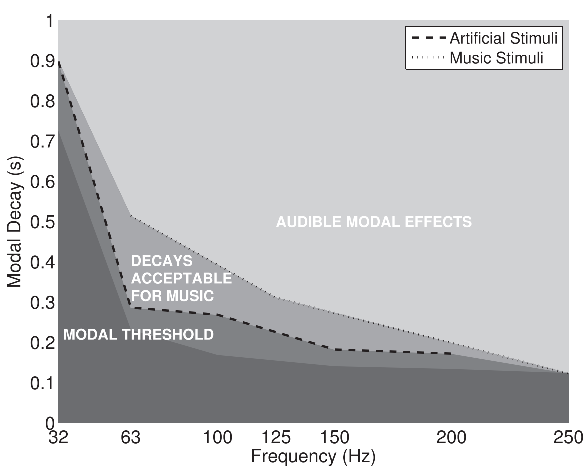

FIG. 7. Map of perceptual regions for modal decay. Modal threshold is defined from experimental results using artificial stimuli at 85 dB. Music threshold is defined from results for both samples at both levels.

VII. CONCLUSION

This paper presented a study on the frequency and level dependency for perceptual thresholds as a function of modal decay. Tests have been carried out with both artificial test stimuli and music stimuli in the presence of a modal sound-field, the latter being more representative of an ecologically valid test regime. Artificial test stimuli were presented in the form of decaying tones in the absence of instantaneous masking. The tests using music stimuli include the presence of instantaneous and non-instantaneous masking. Results show that, in general, perceptual modal thresholds are independent of presentation level, except for thresholds obtained with artificial stimuli below 63 Hz, where a significant effect of level has been found. In this frequency region, the natural rise of hearing sensitivity thresholds combined with post-masking effects appear to play a more dominant role in setting the perceptual thresholds.

The content in music stimuli has an effect on how well modal problems are detected and leads to statistically significant interactions and differences in the thresholds measured. This result indicates that the selection of music samples for testing the perceptual effects of modes is important.

Perceptual thresholds for modal effects when testing with artificial stimuli decrease rapidly with increasing frequency up to about 100 Hz where they appear to level out. For music stimuli, thresholds decrease monotonically with frequency. Average thresholds measured with artificial stimuli are 0.9 s at 32 Hz, 0.3 s at 63 Hz, 0.27 s at 100 Hz, 0.18 s at 150 Hz, and 0.17 s at 200 Hz. Average thresholds measured with music stimuli are 0.51 s at 63 Hz, 0.3 s at 125 Hz, and 0.12 s at 250 Hz. Test conditions using artificial stimuli evoke lower thresholds than those measured with music stimuli given the easier task presented to the subjects and the lack of simultaneous masking presented by other musical events. Thresholds measured with artificial stimuli are therefore considered to provide *worst case* thresholds below which, on average, subjects cannot detect the presence of modal decays, while those measured with music provide a more tolerant working range below which the effects of modes are not detected when listening to typical audio programs.

Results presented here are important in defining perceptually relevant thresholds for measures of modal control and, in general, for further research into aspects of modal perception in critical listening environments, particularly the development of an auditory model.

ACKNOWLEDGMENTS

The authors wish to thank Aki Mäkivirta and Matti Karjalainen for their expert guidance in the earlier work of this study. Additionally, Genelec Oy is thanked for providing facilities during that time. The Universities of Salford and Huddersfield are thanked for their bursaries in support of M.S.'s Ph.D. studies.

[1]R. H. Bolt, "Note on normal frequency statistics for rectangular rooms," J. Acoust. Soc. Am. **18**(1), 130–133 (1946).

[2]M. M. Louden, "Dimension ratios of rectangular rooms with good distribution of eigentones," Acustica **24**(5), 101–104 (1971).

[3]O. J. Bonello, "A new criterion for the distribution of normal room modes," J. Aud. Eng. Soc. **29**(9), 597–606 (1981).

[4]R. Walker, "Optimum dimension ratios for small rooms," in *Proceedings of the 100th Audio Engineering Society Convention* (1996), paper 4191.

[5]T. J. Cox, P. D'Antonio, and M. R. Avis, "Room sizing and optimization at low frequencies," J. Aud. Eng. Soc. **52**(6), 640–651 (2004).

[6]J. A. Pedersen, "Adjusting a loudspeaker to its acoustic environment: The ABC system," in *Proceedings of the 115th Audio Engineering Society Convention*, October (2003), paper 5880.

[7]A. R. Groh, "High-fidelity sound system equalization by analysis of standing waves," J. Aud. Eng. Soc. **22**(10), 795–799 (1974).

[8]R. F. Allison, "The sound field in home listening rooms II," J. Aud. Eng. Soc. **24**(1), 14–19 (1976).

[9]K. O. Ballagh, "Optimum loudspeaker placement near reflecting planes," J. Aud. Eng. Soc. **31**(12), 931–935 (1983).

[10]T. Welti and A. Devantier, "Low-frequency optimization using multiple subwoofers," J. Aud. Eng. Soc. **54**(5), 347 (2006).

[11]B. M. Fazenda, M. Wankling, J. Hargreaves, L. Elmer, and J. Hirst, "Subjective preference of modal control methods in listening rooms," J. Aud. Eng. Soc. **60**(5), 338–349 (2012).

[12]P. Antsalo, M. Karjalainen, A. Mäkivirta, and V. Välimäki, "Comparison of modal equaliser design methods," in *Proceedings of the 114th Audio Engineering Society Convention* (2003), paper 5844.

[13]T. J. Cox and P. D'Antonio, *Acoustic Absorbers and Diffusers: Theory, Design and Application*, 2nd ed. (Taylor and Francis, London, 2009), Chap. 6.

[14]A. Celestinos and S. Nielsen, "Controlled Acoustic Bass System (CABS) a method to achieve uniform sound field distribution at low frequencies in rectangular rooms," J. Aud. Eng. Soc. **56**(11), 915–931 (2008).

[15]A. Goldberg and A. Mäkivirta, "Performance comparison of optimized equalization techniques for in-situ loudspeaker measurements," Proc. Inst. Acoust. **27**(5), 193–205 (2005).

[16]R. Bucklein, "The audibility of frequency response irregularities," J. Aud. Eng. Soc. **29**(3), 126–131 (1981).

[17]F. E. Toole and S. E. Olive, "The modification of timbre by resonances: Perception and measurement," J. Aud. Eng. Soc. **36**(3), 122–141 (1988).

[18]S. E. Olive, P. L. Schuck, J. G. Ryan, S. L. Sally, and M. E. Bonneville, "The detection thresholds of resonances at low frequencies," J. Aud. Eng. Soc. **45**(3), 116–127 (1997).

[19]B. Fazenda, M. R. Avis, and W. J. Davies, "Perception of modal distribution metrics in critical listening spaces-dependence on room aspect ratios," J. Aud. Eng. Soc. **53**(12), 1128–1141 (2005).

[20]M. Karjalainen, P. A. A. Esquef, P. Antsalo, A. Mäkivirta, and V. Välimäki, "Frequency-zooming ARMA modeling of resonant and reverberant systems," J. Aud. Eng. Soc. **50**(12), 1012–1029 (2002).

[21]A. Mäkivirta, P. Antsalo, M. Karjalainen, and V. Välimäki, "Modal equalization of loudspeaker-room responses at low frequencies," J. Aud. Eng. Soc. **51**(5), 324–343 (2003).

[22]M. Karjalainen, E. Piirilä, A. Järvinen, and J. Huopaniemi, "Comparison of loudspeaker equalization methods based on DSP techniques," J. Aud. Eng. Soc. **47**(1/2), 14–31 (1999).

[23]M. Karjalainen, T. Paatero, J. N. Mourjopoulos, and P. D. Hatziantoniou, "About room response equalization and dereverberation," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (2005), pp. 183–186.

[24]M. R. Avis, "IIR Bi-Quad controllers for low frequency acoustic resonance," in *Proceedings of the 111th Audio Engineering Society Convention* (2001), paper 5474.

[25]S. J. Elliott, L. P. Bhatia, F. Deghan, A. H. Fu, M. S. Stewart, and D. W. Wilson, "Practical implementation of low-frequency equalization using adaptive digital filters," J. Aud. Eng. Soc. **42**(12), 988–998(1994).

[26]L. D. Fielder, "Analysis of traditional and reverberation-reducing methods for room equalization," J. Aud. Eng. Soc. **51**, 3–26 (2003).

[27]T. S. Welti and A. Devantier, "In-room low frequency optimization," in *Proceedings of the 115th Audio and Engineering Society Convention* (2003), paper 5942.

[28]M. Karjalainen, P. Antsalo, A. Mäkivirta, and V. Välimäki, "Perception of temporal decay of low-frequency room modes," in *Proceedings of the 116th Audio Engineering Society Convention* (2004), paper 6083.

[29]B. M. Fazenda and M. R. Wankling, "Optimal modal spacing and density for critical listening," in *Proceedings of the 125th Audio Engineering Society Convention* (2008), paper 7584.

[30]M. R. 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. Aud. Eng. Soc. **55**(7–8), 611–622 (2007).

[31]A. Goldberg, "Windowed sine bursts: In search of optimal test signals for detecting the threshold of audibility of temporal decays," in *Proceedings of the 126th Audio Engineering Society Convention* (2009), paper 7812.

[32]A. Goldberg, "A listening test system for measuring the threshold of audibility of temporal decays," Proc. Inst. Acoust. **27**(5), 46–57 (2005).

[33]Lenine, "Meu Amanhã," *Na Pressão*, [Compact Disc], 16–22 s (1999).

[34]H. Cole, "Jersey Girl," *Temptation*, [Compact Disc], 1–8 s (1995).

[35]M. Stephenson, "Assessing the quality of low frequency audio reproduction in critical listening spaces," Doctoral thesis, University of Salford, UK, 2012.

[36]H. Fletcher and W. A. Munson, "Loudness, its definition, measurement and calculation," J. Acoust. Soc. Am. **5**(2), 82–108 (1933).

[37]H. Kuttruff, *Room Acoustics*, 4th ed. (Taylor and Francis, New York, 2000), Chap. 3, pp. 60–63.

[38]M. Wankling, B. Fazenda, and W. J. Davies, "The assessment of low-frequency room acoustic parameters using descriptive analysis," J. Aud. Eng. Soc. **60**(5), 325–337 (2012).

[39]P. M. Morse, *Vibration and Sound*, 1st ed. (McGraw-Hill, New York, 1936), Chap. 8, pp. 306–322.

[40]B. M. Fazenda, "The Perception of Room Modes," Doctoral thesis, University of Salford, UK, 2004.

[41]M. M. Taylor and C. D. Creelman, "PEST: Efficient estimates on probability functions," J. Acoust. Soc. Am. **41**(4), 782–787 (1967).

[42]International Telecommunication Union, "ITU-BS. 1770 Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level," Tech. Rep., International Telecommunication Union (Geneva, Switzerland, 2006).

[43]A. Goldberg, "Measuring the threshold of audibility of temporal decays," in *Proceedings of the 120th Audio Engineering Society Convention* (2006), paper 6823.

[44]E. Zwicker, "Dependence of postmasking on masker duration and its relation to temporal effects in loudness," J. Acoust. Soc. Am. **75**(1), 219–223 (1984).

[45]H. Fastl and E. Zwicker, *Psychoacoustics. Facts and Models*, 3rd ed. (Springer, Berlin, 2007), Chap. 4, p. 78.