# Measurement of sound quality for critical listening

Fazenda, BM, Holland, KR and Newell, PR

Twelfth International Congress
on Sound and Vibration

# MEASUREMENT OF SOUND QUALITY FOR CRITICAL LISTENING

Bruno Fazenda[1], Keith Holland[2], Philip Newell[3]
[1]University of Salford, UK. Technical University of Denmark
[2]ISVR, University of Southampton, UK
[3]Acoustics Consultant, Moaña, Spain
(e-mail address of lead author:) b_fazenda@netcabo.pt

## Abstract

For the critical assessment of music recordings the uniformity of the pressure amplitude response with respect to frequency has robustly been defended as one of the principals of prime importance for any electro-acoustic monitoring system. This has been particularly so in the appraisal of the room responses, where non-linear factors such as intermodulation distortions do not apply. However, despite the traditional emphasis being placed on the even distribution of modal and diffuse energy, recent investigations have begun to indicate that the uniformity of the modal decay rate is far more important than has previously been appreciated in terms of the definition of accurate and revealing monitoring acoustics.

Parallel work on loudspeaker systems has also been tending to lead to similar conclusions. Modulation transfer function related experiments have begun to demonstrate how the effects of temporal response smearing can significantly reduce the ability of any room/loudspeaker system to convey to the ears of the listeners the complex details within musical recordings.

Measurements in real control rooms, and work on the subjective perception of room modes based on binaural room simulations, have now begun to highlight the importance of fast and uniform temporal decays over the whole audio frequency band where monitoring accuracy is a prime requirement. Studies on factors such as modal distribution and decay rates have revealed perception biases that indicate their relative importance in defining perceptually accurate monitoring conditions.

The emerging trend is that to achieve such conditions, the temporal response of electro-acoustic systems must be considered at all times, and the all-pervading pre-eminence of the pressure amplitude response may have to be compromised.

## INTRODUCTION

It has been customary to assess rooms for the critical listening to loudspeaker-reproduced music in terms of the flatness of their steady-state transfer functions – from source to listener(s) – and the uniformity of their decay time when measured in one-third-octave bands. This paper attempts to question inadequate concepts used to describe the quality of reproduction that have found their way into the final specifications for loudspeaker and room performance.

Despite many proposals and years of use, the ‘standard’ listening rooms such as those conforming to IEC or DIN specifications (268-13 and 45 573 respectively) have never achieved the degree of room-to-room compatibility that many people expected. The individual characteristics of the reflexions and resonances give each room a unique ‘fingerprint’ in terms of colouration, and they could not be used, at least not very successfully, to compare loudspeakers of very high definition because they mask much low level detail. However, it must also be noted that many of the IEC or DIN rooms have been constructed and used for purposes for which they were never originally intended, but this has often been done because few other (if any) widely recognised listening room specifications are available. It should also be noted that the IEC and DIN room specifications pre-date the consumer, digital music era, and so were perhaps never intended to deal with the low colouration, low-frequency responses which we can now enjoy.

Due to the effects of resonances (common *room modes*) even state-of-the-art audio control rooms reveal common problems in the low frequency range affecting the performance of its users and hindering sound compatibility and standardisation across this type of facility [1]. Clearly, during the music production process, be it for live performance or for broadcast or recording, the resonances of the performance spaces and any amplification equipment can be regarded as part of the music itself – an extension of the instruments. In fact, the instruments have developed hand-in-hand with any associated performance spaces and amplification equipment. Neither an electric guitar plugged directly into a hi-fi system nor a violin played in a dry acoustic would be considered to be as musically rich as when either are played with the added resonances of electro-acoustic or acoustic ‘help’ to support their tone. Nevertheless, once these composite sounds have been captured by a recording or broadcast system, then it would seem to be a fundamental requirement of any environment in which they will be critically assessed (at least in recording quality terms, even if not in purely musical terms) that it should neither add to nor mask the characteristics of the original recorded sound. The same applies to rooms in which the reproduction equipment is being assessed.

It is becoming ever more evident from recent research that a great failing of the specifications of many listening rooms has been the lack of attention paid to the time responses of the modal energy [2]. Whilst past proposals have concentrated on modal distribution [3,4,5,6] little has been specified about the decay in the modal region. Whilst the benefits of low decay times for music mixing rooms have long been recognised [7,8] there still exists a strong body of opinion in support of longer decay time for the assessment of domestic equipment, but perhaps it may prove to be the case that longer decays and low colouration are not compatible with the requirements if the assessment of fine detail in the sound is also a requirement.

Parallel work on the reproduction quality of loudspeakers has also highlighted problems with the use of low frequency resonance and how it may affect the correct deliverance of program material [9,10,11]. Many types of loudspeaker systems use resonant low-frequency loading to extend the response as far down as practically possible for their box size and intended use – at least in terms of their pressure amplitude responses. As will now be shown, the price which must often be paid for this extension is the loss of low-frequency information content. It may also be reasonably expected that room resonances of high Q may be capable of giving rise to disturbing colouration and loss of detail, even if they do not affect either the flatness of the room response or the one-third-octave decay times to any significant degree.

## THE MODULATION TRANSFER FUNCTION APPLIED TO LOUDSPEAKER LOW-FREQUENCY RESPONSES

In a series of conference papers [9,10,11], it has been shown that the pursuit of ever extended low-frequency responses from ever smaller loudspeakers has given rise to the widespread use of ported enclosures and high-pass electronic protection filters. These loudspeakers do offer a very wide, flat frequency response with reduced risk of damage due to excessive diaphragm displacement, but as shown in [10], at the expense of an accurate time response. It is shown that loudspeakers which do not have such an extended low-frequency response, such as those in sealed enclosures, offer improved transient responses and potentially offer more accurate monitoring of low-frequency detail. Figures 1 and 2 show the on-axis frequency response of two similarly-sized monitor loudspeakers. Loudspeaker A has a sealed enclosure and loudspeaker B has a ported enclosure and uses an electronic high-pass protection filter. It can be seen that loudspeaker B has a more extended low-frequency response than loudspeaker A but that the response rolls off more rapidly. Figures 3 and 4 show waterfall (time / frequency) plots of the response of the two loudspeakers to 4 cycles of a 60Hz sine wave. The response of loudspeaker B is seen to 'ring' for longer than that of loudspeaker A, and, somewhat more worrying, is seen to ring at frequencies other than 60Hz. The subjective implications of this transient behaviour have not been rigorously investigated, but it is known that loudspeakers similar to A are preferred over those similar to B by professionals using loudspeakers for critical monitoring applications [9].

In [11], the modulation transfer function (MTF) was introduced as having the potential to distinguish between those loudspeakers that preserve low-frequency detail and those that lose or scramble that information. The MTF is the basis behind the speech transmission index (STI) measurement system which is widely used to quantify the intelligibility of speech signals passed through acoustic and / or electroacoustic systems. Band-limited, modulated random noise is used as an input signal to the system under test. The depth of the modulation of the system output signal is compared to that of the input signal for a range of noise band frequencies and modulation frequencies. The resultant 'matrix' of MTF values is then weighted and summed to give a single measure of the STI. The STI has been proven a reliable indicator of the degree to which a system loses the information content of a signal, in the case of the STI, through reverberation, background noise and system non-linearity.

Assuming that the system is linear and time-invariant, which is reasonable for a loudspeaker at sensible reproduce levels, the MTF can be estimated from the system impulse response, thus:

$$m(F) = \frac{\int_0^\infty h_f^2(t)\; e^{-j2\pi Ft}\, dt}{\int_0^\infty h_f^2(t)\; dt} \tag{1}$$

Where $h_f$ is the impulse response band-limited with centre-frequency $f$ and $F$ is the modulation frequency. For finite-length, digital impulse responses, this can be written

$$m(F) \approx \frac{\sum_0^N h_f^2(n)\; e^{-j2\pi Fn/F_s}}{\sum_0^N h_f^2(n)} \tag{2}$$

Where $N$ is the total length of the impulse response in samples and $F_s$ is the sample frequency. The level at which the loudspeaker reproduces the different frequencies (as shown in Figure 1) is accounted for by setting an average reproduce level and introducing the limit of audibility (the MAF, minimum audible field) as a form of background noise in each noise band, thus:

$$m(F) \rightarrow m(F) \times \frac{1}{1 + 10^{(L_N - L_S)/10}} \tag{3}$$

Where $L_N$ is the MAF for that frequency band and $L_S$ is the level at which the loudspeaker reproduces that frequency. Figures 5 and 6 show the MTF results for loudspeakers A and B, with each frequency band row of the MTF matrices averaged over modulation frequency and plotted as a form of frequency response with the vertical scale showing a measure of information preservation rather than level. Loudspeaker A, with its more accurate time response, is seen to preserve more low-frequency information than loudspeaker B, despite having a much less extended flat low-frequency response.

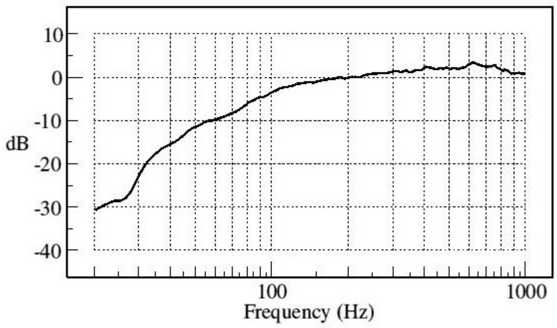


*Figure 1 - On-Axis Frequency Response of Loudspeaker A*

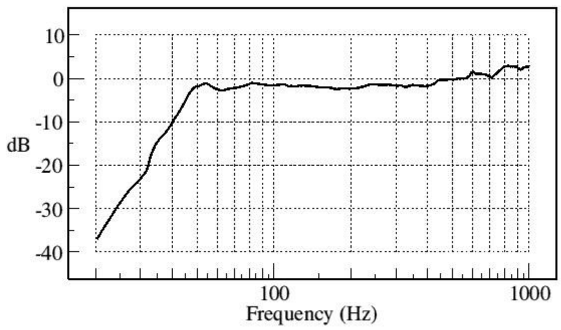


*Figure 2 - On-Axis Frequency Response of Loudspeaker B*

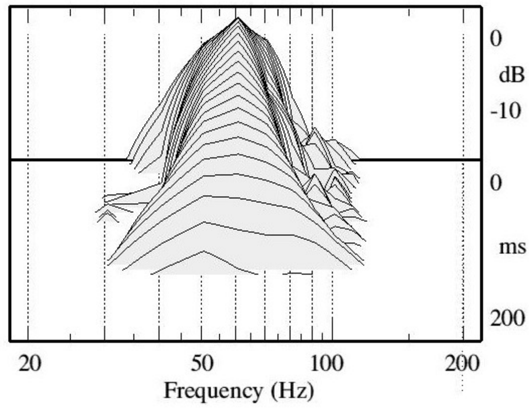


*Figure 3 - Time / Frequency Response of Loudspeaker A to Four Cycles of 60Hz Sine Wave*

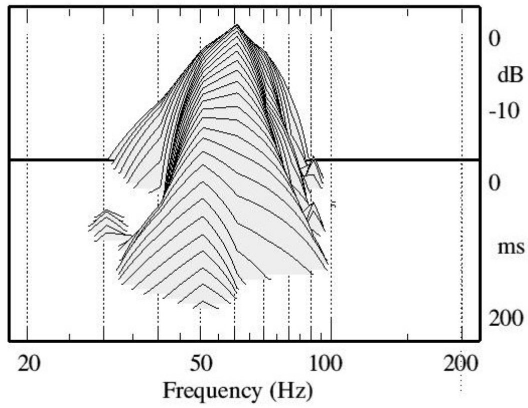


*Figure - 4 Time / Frequency Response of Loudspeaker B to Four Cycles of 60Hz Sine Wave*

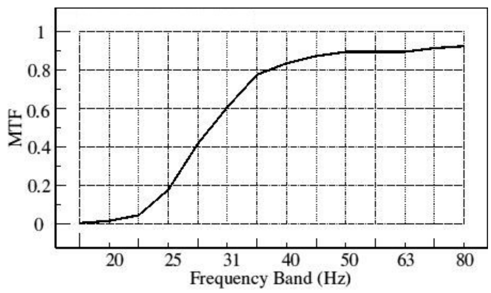


*Figure 5 - MTF Result for Loudspeaker A*

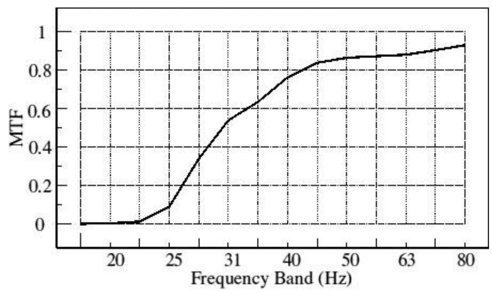


*Figure - 6 MTF Result for Loudspeaker B*

This ability of those loudspeakers with good time responses to preserve information in a signal better than others may go some way to explain their popularity for use in critical listening applications; however, these faster, 'cleaner' responses may be lost in their flight from the loudspeakers to the ears of the listeners if the room modes themselves tend to overwhelm the detail.

## ROOM RESPONSES AND THEIR SUBJECTIVE PERCEPTION

The design of rooms for critical listening has often been based on published recommendations as referred to in the introduction. These recommendations are concerned with avoiding common acoustical problems in the room-transducer-receiver system and attempt to some extent to standardise critical listening conditions.

At mid and high frequencies, the tools available for acoustic treatment are well understood, and adequate control of early reflections or RT is easily implemented. However, at low frequencies, where modal behaviour dominates, adequate control of the acoustic sound field is complex and may involve the use of large, bulky absorption or even active control methods which are still under development and therefore have hitherto had limited application. Perhaps, due to this well known problem and difficulty, the current recommendations 'allow' an increased

RT averaged over the low frequency 1/1 or 1/3 octave bands. Furthermore, there are recommendations for the aspect ratio in an attempt to avoid modal degeneracy and the problems occurring from it. Based on an analysis of the modal frequency distribution directly dependent on the aspect ratios of rooms, authors have published lists or charts indicating the merit of a multitude of possible ratios that may be used to define the outer (rigid) shell of a room [3,5,6,13]. These studies have been based on statistical expressions that describe the optimum distribution of modes as that which would generally be encountered in much larger rooms, like auditoria, where the acoustic conditions at the lower range of the audible spectrum are still considered to be diffuse, and therefore free of the problems of resonance [12].

The methods currently available to rank aspect ratios are all based on the frequency response in the room. Some methods [3] derive a merit score based merely on the frequency spacing between the modes in a given room, and compares this to the 'ideal' modal separation given in [12]. Other methods [6] take into account the number of modes that fall in a given frequency band and establish a list of rules that must be met in order to obtain an 'acceptable' room ratio. The method described in reference [13] uses the magnitude frequency response at some point in the room and compares this to a 'flat' frequency response or some variation to this.

This common reliance on the magnitude of the frequency response appears to be an extension of the common belief that a system that approaches a flat frequency response may be considered of good 'quality' regardless of the time response. Although this has proven sufficient in the description of systems at mid and high frequencies, recent research results show that, at the lower frequencies, the requirement for a flat frequency response may not lead to, and is not sufficient to ensure, accurate monitoring conditions. The previous part of the paper, dealing with loudspeaker responses, has effectively highlighted this point.

Recent studies have been carried out on the subjective perception of changes in the modal distribution of rooms [2]. Results indicate that although changes in room aspect ratios are readily perceived, there appears to be no significant correlation between room ratio 'qualification' according to the above mentioned methods and subjective perception. Furthermore, subjects are able to detect differences, even between rooms that score similarly on any of the metrics studied. It appears that subjects rely on specific stimulus/room 'cues' happening during the progress of the music motif rather than a change in the overall tonal quality between two given room conditions. Consequently, even a room with what may be considered an optimum modal distribution may suffer from problems of resonance every time a note in the input stimulus matches a 'problematic' frequency (resonance or anti-resonance) in the room. These results suggest that the definition of room aspect ratios may still be useful to avoid problems of modal degeneracy, however its reliance on frequency distribution metrics confers a somewhat reduced efficacy to this technique. The logical explanation to these surprising results is that resonance detection may be more closely correlated to other parameters measurable in the acoustic sound field.

Previous work [14,15] on detection of isolated resonances has suggested that the detection of high Q resonances (such as those commonly found in a lightly damped room) may be associated with their decay rates (*ringing)*, which becomes progressively longer with increasing Q and/or decreasing frequency.

Indeed, it can be shown that a high Q resonance system alters the input signal, as shown in Figures 7 and 8 representing the input and output of a single frequency resonant system (Q=20). Input stimulus and resonant frequencies are the same.

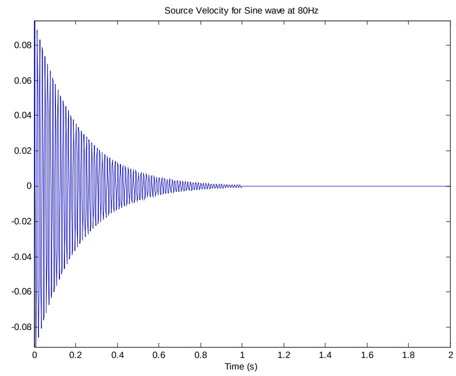


*Figure 7 –Fast Input Stimulus with single frequency*

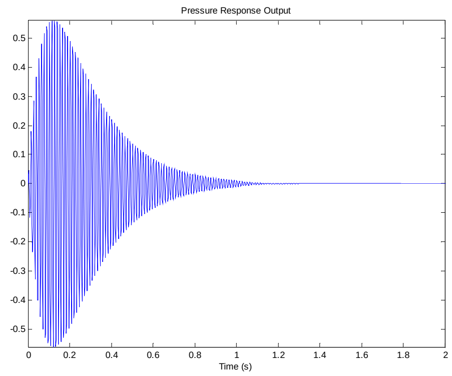


*Figure 8 –Output of Resonant System matching the frequency of the input stimulus*

It is clear that the high Q resonant behaviour causes a temporal 'smear' of the input. The attack and decay portions of the signal are 'distorted'. It is important to recognise an associated gain effect of the resonance – the y scale axes have been normalised.

The response of the same input stimulus in a different system is shown as a waterfall plot in Figure 9. The resonant system is now comprised of two room modes interacting to form an anti-resonance at the stimulus frequency. The figure displays the changes in spectral energy as time progresses from the initial onset of the stimulus to the natural decay stage of the system.

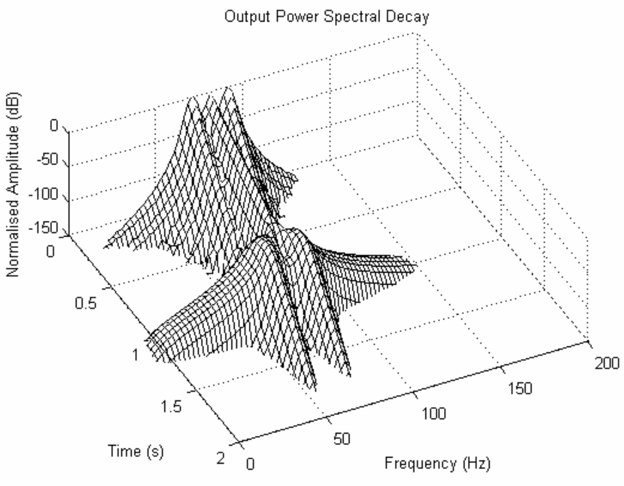


*Figure 9 – Time/Frequency Response for a system with two resonances forming an anti-resonance at the frequency of the input stimulus*

It is clear in Figure 9 that the response of the system exhibits three spectral components at the onset - the stimulus and the resonant frequencies. The effect previously mentioned for ported loudspeaker resonances, is also evident in the transfer of energy to the frequencies of the resonant components occurring during the natural response of the system – when the input is switched off. In an analogous room situation, this stimulus would strongly excite the nearby modes, which could possibly be heard at the onset and during the natural decay of the room. These figures clearly demonstrate that the effects of resonances are not solely imparted to the magnitude frequency response, but also to the time domain where subjective perception may play a very important role in detection of problems.

The audibility of these factors has been subject of further studies [16] and it has been shown that the effects are clearly noticeable for single tones at or near the resonant frequencies, even when magnitude differences between cases have been removed. Results also show that subjects are able to detect differences between resonant systems, albeit only when the spectral energy in the input stimulus is broadband. This is the case for fast transient input stimulus, such as the one represented in Figure 7, very common in most music motifs. This suggests that resonance detection may have a strong temporal dependence as would be expected. Indeed this has been further supported in the same experimental work by results showing that a reduction in the Q-factor of the resonances induces a high degree of uncertainty in their detection. If the decays of the resonances are greatly reduced, the subjects no longer detect their effects on single tones.

In lightly damped rooms the resonances may have very high Q characteristics, and these are known to alter the character of the input stimulus (Figures 7, 8 and 9). It has also been established that these changes are significantly noticed by subjects. Hence, when the subjective quality of an electro-acoustic system is under scrutiny, it is important to take into account the relative importance of the various factors that can be measured in an acoustic sound field. Efforts should then be concentrated on addressing those factors that are more closely related to subjective perception. As a result, it appears more important to design rooms or systems that maintain an adequate temporal ‘transfer’ of the signal being auditioned rather than strive for a constant magnitude frequency response and ignore the subjectively noticeable time effects of resonances.

For the design of a critical listening environment, and according to the concepts presented here, the importance of the temporal response prevails over the magnitude frequency response. The thresholds of detection defined experimentally may be used as guides in the design of the space and its absorption characteristics. If at all necessary, given that a reduction in Q-factor is directly associated with a reduction on peak amplitude of a resonance, any remaining magnitude frequency response discrepancies may be corrected by minimum phase apparatus that should not introduce time distortions of its own.

In such situations, the perception in a listening area is then free of unwanted noticeable effects providing the user with a ‘clearer’ and ‘easier’ environment to work in, and aesthetic decisions will not be hindered by uncertainties in the monitored sound. Furthermore, the compatibility between rooms complying with these new concepts will be higher – given that the individual characteristics of each room are effectively reduced or ‘removed’ – which leads to the much sought ideal of standardisation of listening conditions.

## CONCLUSIONS

Where perceptually accurate listening conditions are required, the findings presented in this paper strongly suggest that short decay rates are a fundamental requirement.

The extreme emphasis which has traditionally been placed on the achievement of the flattest possible pressure amplitude responses would appear not to be justified when the flatness has been gained at the expense of the time response, or when the time response has been assessed only in terms of one-third-octave band energy. Resonances which are long in terms of the average time response of a system (by proportion, rather than in absolute terms) will colour sound and mask low level detail. Whilst the authors fully appreciate that the frequency response magnitude flatness and modal distribution uniformity may be important factors in the responses of rooms which are used for listening enjoyment, it appears that for quality control purposes, either of the recordings themselves or the reproduction equipment, perceptually accurate monitoring requires a high degree of time response accuracy which may only be achieved if the decay in the modal region is close to the threshold of detection.

The traditional thinking regarding room responses has been for uniformly proportional distribution of modal energy, but the findings presented here raise a serious question about whether a neutral perception can be achieved unless modal Q's are well controlled, irrespective of the lack of effect which they may have on one-third-octave pressure distribution. The findings also go some way to explain why the well-promulgated concepts of preferred room ratios have not received the widespread application that they originally appeared to offer. With the incorporation of a new factor of maximum allowable modal Q – perhaps as low as 10 – a much improved degree of room-to-room compatibility would be likely to result, and the use of such rooms for the absolute assessment of domestic sound reproduction equipment would surely be facilitated. It is well known that smooth irregularities in the magnitude of the frequency response can be learned and mentally compensated for, but response inaccuracies in the time domain have proved themselves to be sources of much perceptual confusion.

## REFERENCES

[1] Fazenda, B.M., Davies, W.J. “The views of control room users”, Proceedings of the Institute of Acoustics, Vol 23, Pt 8, 2001

[2] Fazenda, B.M., Avis, M.R., Davies, W.J., Jacobsen, F “Perception of modal distribution in critical listening spaces”, Proceedings of the 11th International Conference on Sound and Vibration, July 2004, St. Petersburg, Russia

[3] Louden, M. M. “Dimension-Ratios of rectangular rooms with good distribution of eigentones”, Acustica, Vol. 24, 1971

[4] Gilford, C.L.S., “The acoustic design of talks studios and listening rooms”, Journal of the Audio Engineering Society, Vol. 27, No. 1/2, Jan/Feb 1979

[5] Bonello, O. "A new criterion for the distribution of normal room modes." Journal of the Audio Engineering Society, Vol. 29, No.9, pp.597-605, 1981

[6] Walker, R. “Low Frequency room responses – Part 2 – Calculation methods and experimental results”, BBC RD 1992/9

[7] Newell, P.R., Holland, K. R., Hidley, T., “Control room reverberation is unwanted noise”, Proceedings of the Institute of Acoustics, Reproduced Sound 10,Vol.16, Part 4, pp. 365-373, Windermere, UK, 1994

[8] Wrightson, J. “Psychoacoustic considerations in the design of control rooms”, Journal of the Audio Engineering Society, Vol. 34, No. 10, October 1986

[9] Newell, P.R., Holland, K.R. Newell, J. P., “The Yamaha NS10M: twenty years a reference monitor. Why?”, Proceedings of the Institute of Acoustics, Reproduced Sound 17, Vol.23, Part 8, pp. 29-40, Stratford-U-Avon, UK, 2001

[10] Holland, K. R., Newell, P., Mapp, P. “Steady state and transient loudspeaker frequency responses”, Proceedings of the Institute of Acoustics, Reproduced Sound 19, Vol.25, Part 8, Oxford, UK, 2003

[11] Holland, K. R., Newell, P.R., Mapp, P. “Modulation depth as a measure of loudspeaker low-frequency performance”, Proceedings of the Institute of Acoustics, Reproduced Sound 20, Vol.26, Part 8, pp. 107-115, Oxford, UK, 2004

[12] Bolt, R. H. “Normal modes of vibration in room acoustics: angular distribution theory”, Journal of the Acoustical Society of America, Vol. 11, p.74, 1939

[13] Cox, T.J., D'Antonio, P., Avis, M.R. “Room sizing and optimization at low frequencies” Journal of The Audio Engineering Society 52 (6): 640-651 JUN 2004

[14] Bucklein, R. “The audibility of Frequency Response irregularities”, Journal of the Audio Engineering Society, Vol. 29, No. 3, March 1981

[15] Olive, S., P. Schuck, Ryan, J. G., Sally, S. L., Bonneville, M. E. "The Detection Thresholds of Resonances at Low Frequencies.", Journal of the Audio Engineering Society, Vol. 45, No. 3, March 1997

[16] Fazenda, B.M., Avis, M.R. “Perception of low frequencies in small rooms”, Proceedings of the European Acoustics Symposium, September 2004, Guimarães, Portugal