A LISTENING TEST SYSTEM FOR MEASURING THE THRESHOLD OF AUDIBILITY OF TEMPORAL DECAYS

A. Goldberg Genelec Oy, Olvitie 5, 74100 IISALMI, Finland.

1 INTRODUCTION

The aim of this related series of studies is to gain a greater understanding of the sensitivity of human hearing to temporal smearing (longer decay times at certain frequencies due to room or other system resonances) particularly at low frequencies. The hypothesis is that the absolute hearing threshold limits the human capacity to recognise both narrow-band resonances/antiresonances and temporal smearing at low frequencies, and secondly that the sensitivity to hear either magnitude response errors or temporal smearing is further decreased in real room conditions.

The study series firstly describes the design of a suitable listening test for identifying the audibility of resonances, secondly conducts the listening test and analyses the results, and finally tests existing audibility models for applicability to the listening test results. If existing models prove to be inadequate, modifications will be proposed. Initially, the aim is to understand how the human auditory system performs in this respect under ideal listening conditions (headphone listening) by eliminating the room as a variable. Later research may seek to understand how the auditory system performs in typical room listening conditions.

In earlier investigations, upward deviations in the magnitude response have been shown to be more audible than downward ones [1]. It has also been demonstrated that signals with continuous spectra, such as white noise, are more revealing of resonances than signals having discrete spectra, such as solo instrument music, and can be explained by a requirement for the test signal to excite the resonant frequency for it to be audible. Even so, listeners needed extensive training to learn to detect resonances. With the exception of two test frequencies at 85 Hz and 150 Hz, the resonant frequencies in these tests were located above 200 Hz.

Further research on resonance audibility [2] investigated 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. Amongst other things, the study demonstrated that temporal changes and reverberation time affect the threshold of detection of resonances. Again, the tests concentrated on frequencies above 200 Hz.

Investigations into lower frequency resonances have been made down to 63 Hz [3], but with a relatively sparse frequency resolution; 63, 125, 250, and 500 Hz. The Up-Down Transformed Response Rule (UDTR) was employed to set the test level in a 2AFC listening test. These findings were based on only four listeners, three of which had direct affiliation to the paper's content, and an attempt to fit the results to existing auditory models [4] resulted in major inconsistencies.

The audibility of multiple resonances [5] was investigated by modelling a room down to 34 Hz using biquad filters and replaying sound samples through headphones. The Q-factor of all the room modes were simultaneously varied in accordance with PEST rules [6] and the difference limen deduced. The motivation for this was to ascertain what room ratios are preferable for optimal modal distribution [7].

Another approach for improving the listening experience involved optimising the number, location, and responses of multiple subwoofers [8], such that seat-to-seat frequency response variation is minimised. Equalisation with a somewhat coarse resolution was employed. An investigation into three prevalent equalisation techniques for in-situ loudspeaker equalisation [9] showed that most of the required equalisation is seen at frequencies below 400 Hz. This is due to factors such as boundary loading, reflections causing comb filtering, and room modes.

More sophisticated equalisation techniques seek to invert the frequency response to achieve a close approximation to a unity transfer function (no change to magnitude or phase) within a certain bandwidth of interest [10 - 13]. These techniques lead to side effects in the temporal response, which may, or may not, be beneficial.

Focusing more on the time domain, modal equalisation has recently grown in popularity as a research topic for reducing modal decay time [14 - 16]. The decay time is identified [17, 18] and a modal equalisation technique reduces the pole radii of the modes in the overall transfer function [16]. An alternative method finds peaks in the low frequency response, assumes they are due to resonances and introduces notch filters to reduce the decay time [19].

Whilst objective improvements are readily quantified, recent subjective listening tests suggest that below 100 Hz the threshold of audibility of temporal decay time is high [20]. Those listening tests were conducted at low listening levels and with a sparse frequency resolution, so the question raised is, what is the threshold of audibility of resonances with respect to frequency, level, signal type, and the listening environment? In addition, closely spaced resonances can cause additional effects such as beating in the overall decay pattern, which may alter the audibility threshold.

In summary, objective system equalisation is no longer the limiting factor in improving the listening experience, so the focus has moved towards a listener's ability to perceive an audible sound quality improvement [11]. The subjective research to date shows that the sensitivity of hearing at frequencies below 200 Hz to narrow-band resonances and notches (changes in magnitude response level) or to narrow-band decay (a lengthening of the room decay time at resonant frequencies) has not been studied sufficiently to precisely understand the capability of the human hearing system to these two factors.

This paper opens with a method for generating artificial room modes in audio signals. Possible test signals are classified prior to discussing their reproduction. A description of two methods to ascertain the threshold in the psychometric function is followed by a short description of the listening test system. Recommendations for the full listening test follow a test run of the system to examine its effectiveness.

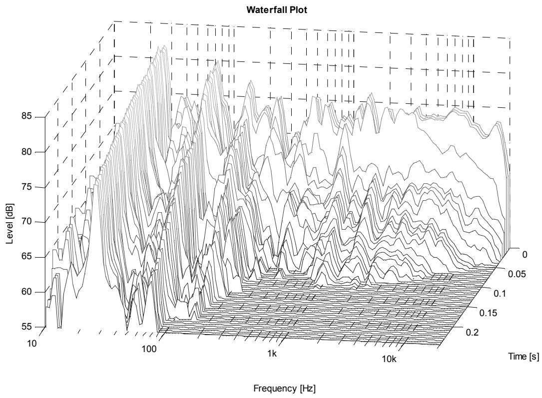
2 TEMPORAL DECAYS AND TEST SIGNALS

2.1 Naturally Occurring Temporal Decays

Resonances in rooms (room modes) are damped and therefore exhibit a decay rate, hence the term temporal decay. Damped resonances with extended decay times are prevalent in listening environments, even in high quality listening rooms. There are often one or more dominant room modes, which are visualised in a waterfall plot or reverberation time curve (Figure 1). It can be seen in this case that a 1/3 octave band T_{60} curve has insufficient resolution to show the fine structure of the temporal decay.

2.2 Test Signals

Test signals are designed to reveal the attribute under study, which in this case is the audibility of a temporal decay. Masking of this response aberration by the test signal itself should be minimised so as to reveal the minimum threshold.



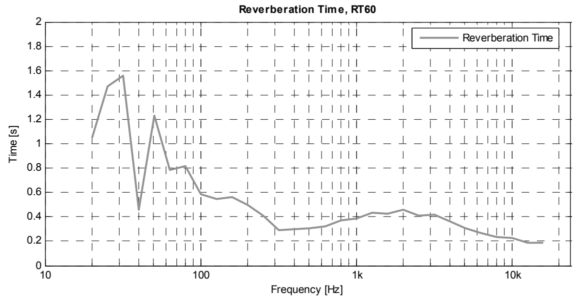


Figure 1. A measurement of a room with insufficient low frequency acoustical treatment. Left: A third-octave smoothed cumulative spectral decay showing strong room modes at 25 Hz and 75 Hz. Right: Third-octave reverberation time of the same measurement.

2.2.1 Artificial Test Signals

In this research, artificial signals are defined as those generated by a computer, such as pure tones, sinusoidal sweeps, and noise. Linear sweeps linger excessively in the low frequency region, which forward masks [4] the temporal decay requiring detection and leads to higher than minimum threshold values. A more effective signal is a fast (<1 s) logarithmic sweep (20 – 1000 Hz), which reduces forward masking thereby revealing a lower threshold value. Interestingly, sweeps were not used in the listening tests cited earlier in this paper.

Noise has previously been identified as a sensitive test signal for revealing resonance thresholds [1 - 3]. Noise used in listening tests typically has a white or pink (-3 dB/octave filtered white noise [21]) spectrum, and can be wideband or narrowband. Steady state resonances (upward deviations in the magnitude response) are easily added to noise signals however, some care is required when introducing temporal decays to noise. For example, should the temporal decay occur during the noise signal or at the end? Concurrent presentation of the masking and probe (temporal decay) signals reveals the simultaneous masking threshold, whereas a probe signal added to notched masking signal or after a masking signal yields the non-simultaneous masking threshold [4]. The threshold in these two cases may be different. Multitone or gated sine burst signals may also prove to be interesting.

2.2.2 Natural Test Signals

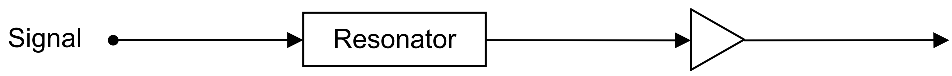
In these studies, natural signals are classified as those not described above and include music, speech, and movie sounds or special effects. An issue with natural sounds in controlled listening tests is that the signal must have content at the frequency of interest for it to excite the resonance, for example, speech is of little use for room mode studies below 100-150 Hz. These temporally and spectrally complex signals may yield higher decay threshold values than those seen for simple artificial signals due to time- and frequency-domain masking.

2.3 Synthetic Temporal Decays

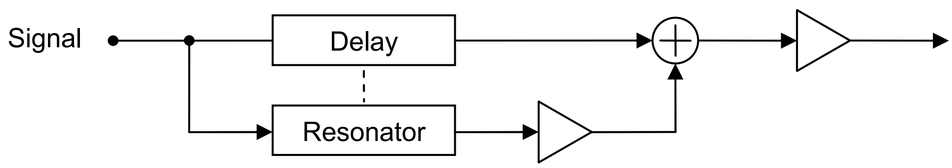
2.3.1 Topology

Three methods to generate a temporal decay in a signal using resonators are shown below. Care must be taken in the Type B and C implementations to guarantee continuity and not to change the magnitude response.

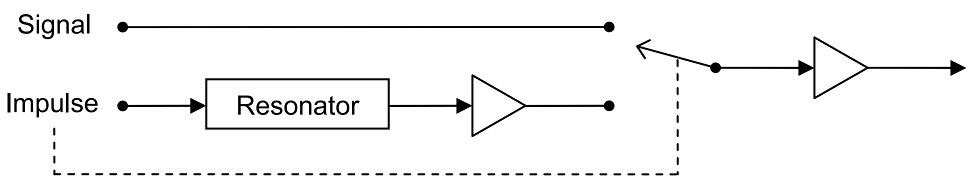
- **Type A**
  The input signal is fed through a resonator and the gain stage is used to set SPL. This topology is very simple to implement.



- **Type B**
  The input signal is fed through a resonator and then added to the delayed original signal. The delay stage compensates for the average delay of the resonator. The gain stage in the allpass resonator loop allows for mixing in the signal summing block and the final gain stage sets SPL. This is a problematic topology as multiple parameters require adjustment and the resonator delay is not linear with frequency so summing is not accurate.



- **Type C**
  An impulse is fed through a resonator and a switch determines when the signal or an impulse generated temporal decay is heard. The final gain stage sets SPL but resonator gain scaling may be difficult depending on the signal.



2.3.2 Resonator Filter

A temporal decay may be simulated in the z-domain transfer function by an allpass complex conjugate pole pair with poles on the inside and zeros on the outside of the unit circle,

$$H_{\text{mode}}(z) = \frac{\left(1 + re^{j\theta}z^{-1}\right)\left(1 + re^{-j\theta}z^{-1}\right)}{\left(1 - re^{j\theta}z^{-1}\right)\left(1 - re^{-j\theta}z^{-1}\right)} \tag{1}$$

where r, the pole radius, is defined by the decay time, T_{60},

$$r = e^{\frac{(\log 0.001)/T_{60}}{f_S}} \tag{2}$$

and θ, the pole angle, is defined by the resonance frequency, f_{res},

$$\theta = \frac{2\pi f_{\text{res}}}{f_S} \tag{3}$$

The allpass filter implementation has unity gain but suffers from a 270° phase change at points where the signal exhibits a transient. An alternative method for simulating a temporal decay is a minimum phase version of the above (same pole positions but zeros moved to the origin), where there is no phase change at a transient but a gain that requires compensation. This is sometimes known as an all-pole allpass filter, which technically it cannot be as it has a non-unity transfer function.

Figure 2 shows two examples of synthetic temporal decays added to artificial test signals using the Type A and Type C methods. To keep the implementation simple for the trial listening test in this paper, an allpass filter is used in the Type A topology.

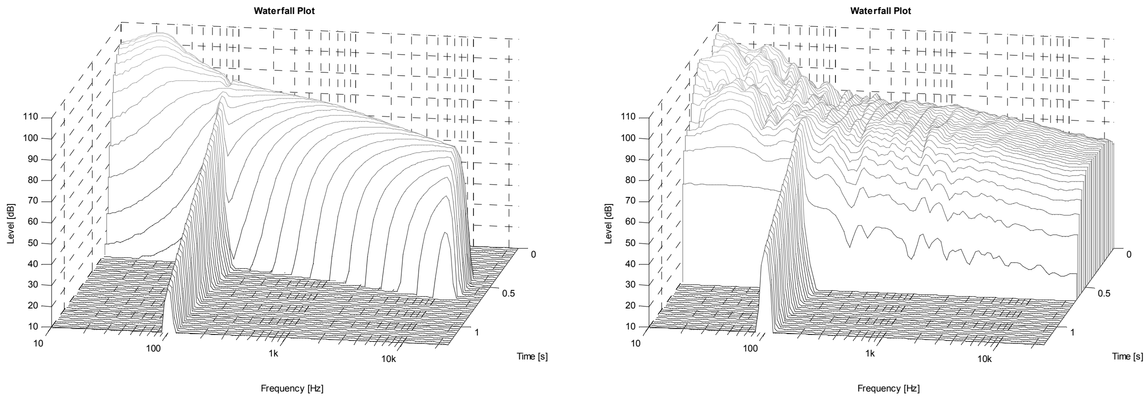


Figure 2. Left: Using the Type A temporal decay generation method, 1/3 octave smoothed cumulative spectral decay of a 0.5 second logarithmic sine sweep (20 – 20000 Hz). It is headphone response equalised with a 1 second temporal decay at 100 Hz added. Right: The same but for 0.5 seconds of pink noise and the Type C temporal decay generation method.

3 TRANSDUCER AND SYSTEM CALIBRATION

Loudspeakers used in listening tests introduce environmental factors such as boundary loading, reflections, reverberation, and noise. Headphones minimise the effect of these factors. The selected headphones are a circum-aural design (Sennheiser HD590) and the response was measured using a dummy head, a calibrated laboratory grade measurement microphone and pre-amplifier, and acoustic measurement software. The response of the headphones is smooth up to 1 kHz [22, 23], so a 3000-tap FIR filter is used flatten the magnitude response of each earpiece up to 1 kHz. In the time domain, this transducer is well behaved, other than for a resonance of 0.15 s (-60 dB) at 90 Hz. Distortion measurements of the headphones show that, at 100 Hz, 2nd order harmonic distortion is <2%, and 3rd order harmonic distortion is <0.5% for reproduction levels up to 110 dB SPL. At 20 Hz, 2nd order harmonic distortion is <1% up to 85 dB SPL and 3rd order harmonic distortion is <0.5% for reproduction levels up to 85 dB SPL.

A 24-bit soundcard with balanced-line outputs (Digigram VXpocket) and a custom-made high quality headphone amplifier (nominal gain 10 dB) drive the headphones. Self-generated noise is sufficiently low not to disturb the listener. The soundcard and amplifier were not calibrated for their complex frequency response as this was considered an insignificant source of replay error. After calibration, the listening test system's replay level is accurate to better than ±0.1 dB at 1 kHz and 94 dB SPL.

4 MEASURING THE TEMPORAL DECAY RATE THRESHOLD

4.1 The Psychometric Function

The psychometric function $P_C(x)$ relates the probability of a correct response P_C to the level of the physical stimulus under study x, with a threshold θ, (Figure 3). In general terms, it is,

$$P_C(x) = \gamma + (1 - \gamma - \lambda)F(x;\theta,\sigma) \qquad (4)$$

where θ is the stimulus threshold based on a probability figure depending on the test methodology or it is taken as the maximum of the slope [24]. γ is the guessing rate (chance performance) which is a function of the task design and therefore in the control of the experimenter. λ is the lapse rate which should have negligible influence in a well-designed experiment as it is a measure of the listener's concentration and fatigue. σ is a spread parameter (inverse slope) which is a measure of reliability of the threshold measurement and the system under test. $F(x;\theta,\sigma)$ can be almost any sigmoid function with asymptotes at 0 and 1 [25], such as the logistic, Weibull, Quick, cumulative normal, cumulative Poisson, hyperbolic tangent [25], and arctan [20] functions. Logarithmic and linear x-axis scales are both commonly used and a z-score conversion is occasionally employed [26]. In this research the Weibull function is fitted to unconverted data using a linear axis,

$$F(x;\theta,\sigma) = 1 - e^{-e^{(\sigma(x-\theta))}} \tag{5}$$

Often one or more of the parameters are fixed at some arbitrary value and the remaining parameters optimised to fit the chosen sigmoid function to the data. In this case, a nonlinear least squares algorithm based on the interior-reflective Newton method (`lsqcurvefit` [28]) fits the parameters θ, λ, γ, and σ of Eq. (4) and Eq. (5) to the acquired listening test data $P'_C(x)$ such that,

$$\min_{\theta,\sigma,\gamma,\lambda} = \sum \left[P_C(x;\theta,\sigma,\gamma,\lambda) - P'_C(x)\right]^2 \tag{6}$$

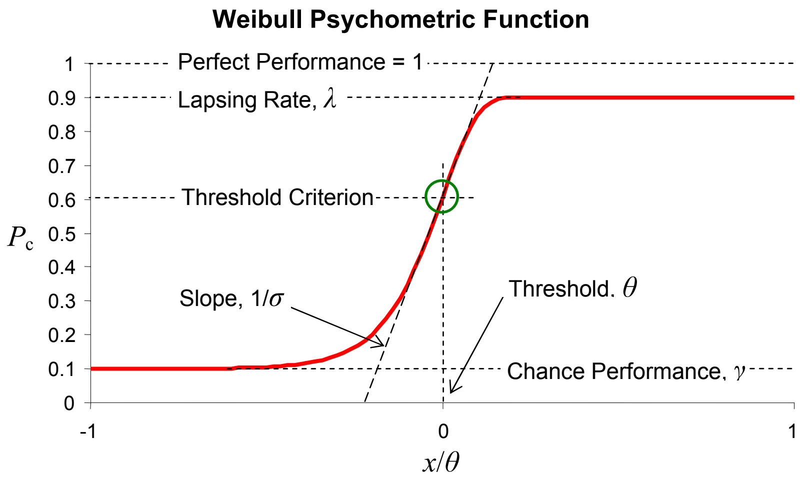


Figure 3. The psychometric function modelled with a Weibull function (parameters θ, λ, γ, σ).

4.2 Measuring the Psychometric Function

There are many ways to measure the stimulus threshold such as Parameter Estimation by Sequential Testing (PEST) [6, 29], Maximum Likelihood Method (ML) [29], Adaptive Two Alternative Forced Choice (2AFC) [30], Fixed Sequence 2AFC [30], and Method of Adjustment (MOA) [30]. These test methods either directly output the threshold value or produce data to which a psychometric function is fitted [25, 27, 31], but some factors serve to complicate this process; guessing rate, lapsing rate, threshold definition, threshold criterion, and slope bias. In this research, the number of variables is large (program material, replay level (SPL), resonant frequency, decay time) so an efficient test method that directly identifies the threshold as a function of a trial's fixed variables is preferable.

PEST is an adaptive procedure designed to determine the threshold of the stimulus both rapidly and efficiently [6, 32]. It is maximally efficient for trial-by-trial sequential decisions at each stimulus level so there is convergence towards a target level, in this case the threshold of audibility of the decay time of a damped resonance. However, it does assume some prior knowledge of the psychometric function. The first stimulus (decay time) level is normally set to be easily discriminated to help the listener identify the feature (resonant frequency) on which they should be focusing. This level does not affect the result as the stepping rules quickly steer the level towards the threshold. On each trial, the listeners choose between two samples, one with a temporal decay, and one without. At each stimulus level, a Wald sequential likelihood ratio test [33] determines whether the stimulus level remains unchanged, increases or decreases. Stepping rules define the step size of the stimulus level:

1. On every reversal of step direction, halve the step size.
2. The second step in a given direction, if called for, is the same as the first.
3. The forth and subsequent steps in a given direction are each double their predecessor (maximum 4 dB change).
4. Whether a third successive step in a given direction is the same as or double the second depends on the sequence of steps leading to the most recent reversal. If the step immediately preceding that reversal resulted from a doubling, then the third step is not double, while if the step leading to the most recent reversal was not the result of a doubling, then this third step is double the second (maximum 4 dB change).

The PEST run is terminated when the defined minimum step size is reached. This minimum step size determines final precision, not efficiency. The threshold of audibility of the decay time emerges as the output of the PEST run. A reliable listener should give consistent results on hidden controls in the listening test otherwise there is reason to remove a listener's results.

Although PEST is efficient and a good choice for determining the stimulus threshold, many trials are required, leading to long testing times, possible listener fatigue, and hence listener lapses. Another shortcoming of PEST is the use of inappropriate starting levels and final step sizes. Since PEST was originally defined, several modifications have been proposed (example [34]), most of which have since been rejected for various reasons [35].

Fitting a psychometric function to the PEST run data may quantify other parameter values [27, 25, 31, 36]. Slope σ informs how gradual the transition is from no detection of the temporal decay to guaranteed detection. Slope estimates are often biased high with the PEST methodology, although this can be corrected depending on the number of trials [36]. Lapsing λ and chance γ performance rates give an indication of the quality of a listener's results and the experiment design respectively. Fitting a psychometric function to PEST data can be problematic due to the distribution of the measured data points.

4.3 Listening Test System

The listening test system (Figure 4) is written entirely in Matlab [37] and runs on a PC. Each listener uses a unique fully factored randomised listening test design. Extensive instructions and practice test runs familiarise the listeners with the task. The software instructs the listener to take breaks at regular intervals to reduce listener fatigue. A questionnaire alerts the test organiser of possible issues that would reduce the ability of a listener to conduct the test at full effectiveness and with a performance similar to the other listeners.

5 TRIAL LISTENING TEST

A trial test was conducted as an assessment of the listening test system efficacy in determining the threshold of audibility of temporal decays. Large-scale listening tests to gain a fuller understanding of temporal decay thresholds follow at a later date.

5.1 Trial Listening Test Design

The factors selected for this listening test are single temporal decays at 32, 50, 80, 125 and 200 Hz, at signal replay levels of 70 and 85 dB SPL. 100 Hz controls replayed at 70 dB SPL are used to check the listeners' performance. These factors are fixed for each PEST run and the decay time (stimulus level) is adjusted for each trial according to PEST rules. The signal is a 10-1000 Hz, 0.5 s, upwards log sweep and the starting decay rate is based on previous informal tests. A minimum step size based on the required accuracy and number of listeners is used to determine the threshold. The resonator used is an allpass filter implemented in the Type A topology.

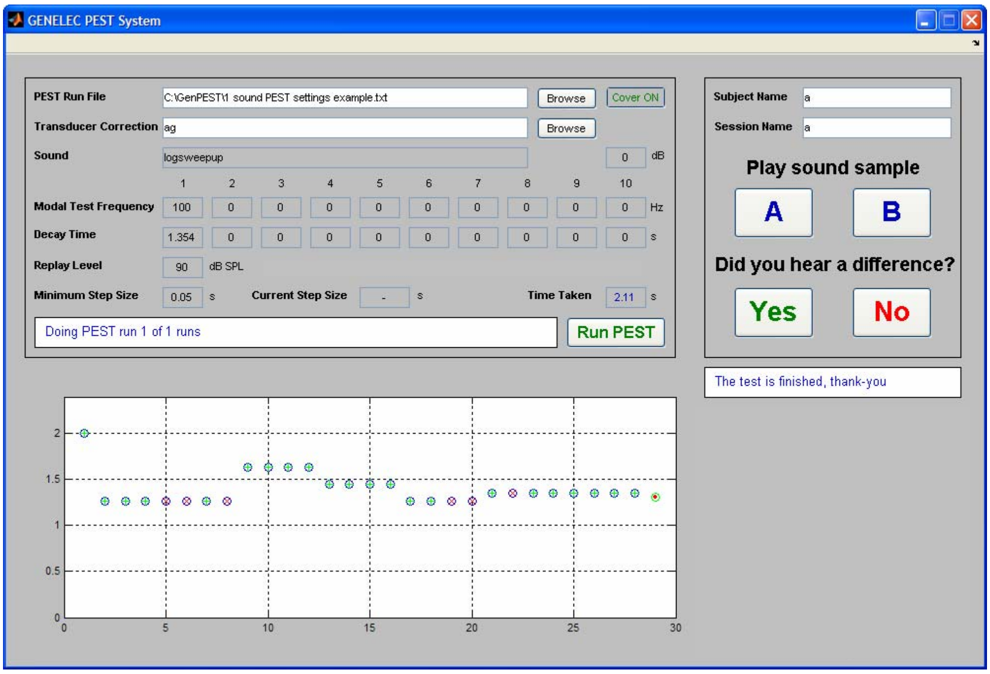


Figure 4. Listening test system graphical user interface. The left side is covered during the listening test to avoid listener bias. A ○ indicates the stimulus level being tested. A + indicates a positive response (audible difference between A & B). An x indicates a negative response (no audible difference between A & B). The ◉ indicates the result, which in this example is 1.354 s.

5.2 Trial Listening Test Results

Six listeners with a median and standard deviation age of 32 ± 13 years, are included in this trial run. Two of the listeners, who are associated with this research program (but not the author), have participated in other listening tests, and all listeners can be described as critical listeners. The least experienced listener reported some known noise induced hearing loss and coincidentally gave the least consistent results. Inconsistent results were also seen for a listener who reported an additional “delay” cue in the test signal during the listening test. In all but one case (the least experienced listener described above), the listeners submitted lower values for the second control signal than the first (100 Hz at 70 dB SPL). The listening tests were conducted in an acoustically well-treated room (low background noise level, T_{60} = 0.2 s) and took 84 ± 17 minutes, with approximately another 30 minutes for listener instruction and two practice PEST runs. The median number of trials in each PEST run was 42 ± 13 and the median number of trials per listening test was 479 ± 80. The number of times each sample (A and B) was played in a trial was not recorded, however it was observed that some listeners were faster in their decision making than others.

To build up the threshold of audibility curves as a function of resonance frequency, SPL, and signal type, the data for each listener is pooled into the factor categories. Taking the median of the pooled results reveals the threshold of audibility (Figure 5 left) and standard deviation error bars show the data spread. The general pattern is that lower frequency temporal decays have a higher threshold and spread of results, however above 80 Hz SPL appears not to be a factor. The standardised standard deviation (STD std dev) shows increased data spread at higher frequencies. In addition, the lower SPL results show a higher threshold at lower frequencies. Fitting a Weibull function to the PEST data (Figure 6) shows similar results except at 32 and 50 Hz where the curve fitting fails at both SPL’s (Figure 5 right). Good intra-listener consistency is seen down to 100 Hz at both SPL’s but at lower frequencies there is an increased spread of threshold values. Even with an initial stimulus level of 3 s, half the listeners could not hear the 32 Hz resonance in the 70 dB SPL signal. Some listeners noted audible distortion for the lower frequency decays, which is function of the transducer. Other listeners reported a cue they called “delay before the ringing”, which was later traced to the initial phase of the temporal decay. In some cases, listener performance changed during the test when this additional cue was perceived.

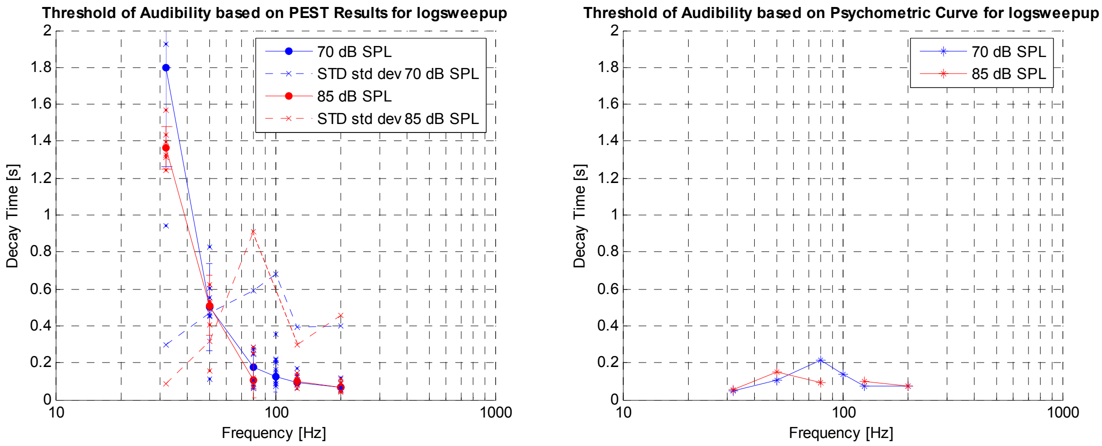


Figure 5 – Threshold of audibility at different SPL's derived from the PEST output (left) and threshold derived from a psychometric function fitted to all the listener's responses pooled together (right).

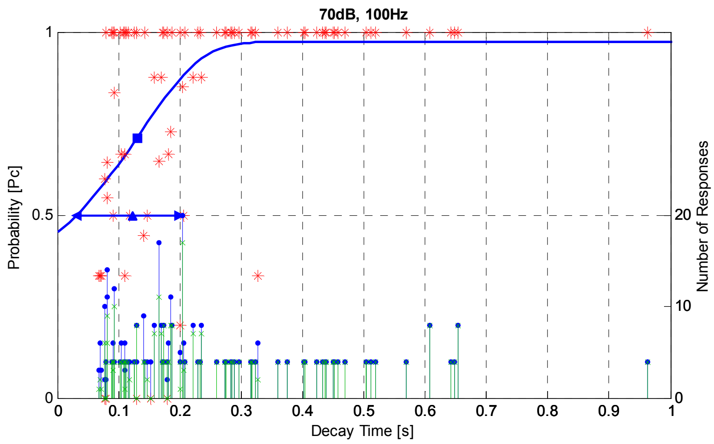


Figure 6 – A Weibull function with parameters θ, λ, γ, σ fitted to the PEST data for listener's responses for frequency and SPL pooled together: total number of responses (dots), number of "*yes*" responses (crosses), percentage "*yes*" responses (stars), modelled psychometric function (line), psychometric function threshold (square) and PEST threshold with standard deviation (triangle and arrowed line).

6 DISCUSSION

6.1 Listening Test System Design and Conditions

The listening test system described above works as realistic results are obtained with experienced listeners. In fact, it is sensitive enough to reveal deficiencies in the signal-resonator combination – see later comments. As has been observed in many other listening tests, practice improves listener performance and reduces listening test time. There is objective evidence, in the form of lower results for the second control signal, that the listeners' audibility performance improved during the test. This indicates that more careful listener selection and a longer training period are required. In addition, despite the use of mandatory breaks, there is some anecdotal and objective evidence of listener fatigue due to the test length, so it appears that 1 hour and breaks every 20 minutes is the maximum tolerable test session time.

The initial stimulus level (temporal decay time) does not affect the threshold result and so the decay time may be reduced. However, it must long enough to be clearly audible at a lower SPL for low frequency decays, but not so long that the PEST run length becomes unnecessarily long. Results show that initial decay times at higher frequencies can be significantly shortened to reduce the overall listening test time. Termination step sizes should be reviewed for the same reason, however this affects accuracy of the results. Listener lapses due to inaudibility of the resonance lengthen the test time, but, as the PEST adaptive algorithm is robust to lapses, it recovers quickly.

The headphones are an open-back design as so exhibit high out-to-in sound leakage. Although not tested formally, there is some evidence that listening room background noise may bias results. Impulsive noise probably has little effect, because the listener may replay the sounds as required, however tonal noise may influence the results and should be minimised.

6.2 Test Signal and Temporal Decay Modelling

It is not desirable to have audible non-modal cues. One such cue is distortion at low frequencies, which some listeners reported. This distortion can be reduced by increasing the sweep speed, starting the sweep at a higher frequency, and/or reducing the SPL. Another audible non-modal cue comes from the initial phase of the resonator being non-zero (270°). When the signal excites the resonance, the temporal decay starts with a negative step and so it interferes with the signal that is still sweeping upwards in frequency. This anomaly explains most of the spread in threshold values as half the listeners perceived a “delay before the ringing” or “click” whereas the others did not, thereby creating two populations of results at lower frequencies. This phase problem in the decay onset must be addressed to remove the audible non-modal cue and hence give more consistent results. An all-pole minimum-phase resonator configured in various topologies is currently under consideration for creating the modal resonance in the test signal. Other test signals will be required for the large-scale listening test.

6.3 Threshold Values

The results indicate that the threshold of the decay time of a temporal decay is higher at lower frequencies. Currently it is suspected that this is related to the increase in absolute hearing threshold with lower frequency, but this will be tested later in the program of studies. Thresholds may also be higher for lower SPL, however fixing the problem with the initial phase of the temporal resonance filter should yield a more definitive answer. Many rooms, including those used for critical listening, have decay times in excess of the audibility thresholds measured in this trial listening test, especially at higher frequencies. Rooms may start to “sing” when sound sources are loud because the room modes become audible, so modal equalisation is probably required for sound systems capable of high SPL, especially if installed in smaller rooms suffering from room modes above 80 Hz.

There is a knee point in both curves below 80 Hz (Figure 5 left) so an increased density of resonant frequencies is required. The addition of 63 Hz and 40 Hz gives a one-third octave resolution below 80 Hz and a two-thirds octave resolution above 80 Hz. Adding a 400 Hz frequency point extends the data set into the mid band to include higher frequency resonances observed in some listening rooms and to allow comparison against other research which mainly focuses on this region.

Modelling the psychometric function (Figure 5 right), especially at low frequency, was either unsuccessful or inaccurate as the PEST data has insufficient values near the origin. This especially affects the chance performance parameter and consequently the slope and threshold parameters. Compounding the problem is the listeners’ data falling into two distinct populations: those that noticed “delay before the ringing” and those that did not. Improved listener training and modal decay modelling should remove this problem. The modelling was successful at higher frequencies as the spread of listener data was narrow.

The thresholds reported here are low compared to recent research [20] as the test signal with a single resonance is artificial and analytical. Natural signals with complex time and frequency variance may yield higher values due to time- and frequency-domain masking. Conversely, closely spaced multiple resonances can result in beating in the decay curve. This additional cue may lead to lower threshold values. Further listening tests with a variety of signals are required to build a complete low frequency psychoacoustical model.

7 CONCLUSIONS

A listening test system for discovering the audibility of slowly decaying room modes at low frequencies has been described. It is based on the PEST methodology. Headphone equalisation is included, as is the provision for artificial and natural test signals. Artificial temporal decays are added to the test signal and the decay time varied until the threshold of audibility is discovered using the adaptive algorithm. A small-scale listening test shows that the system yields realistic values although, as this was only a trial run, this has not been checked against the literature. More consistent results are expected with increased listener practice and improved temporal decay modelling in the signal processing. The trial listening test revealed deficiencies in the temporal decay model, so additional work is required before commencing large-scale listening tests. The listening test system itself was effective in finding the threshold of audibility of temporal decays and is ready for large-scale listening tests to gather data for testing temporal decay audibility models. Early indications are that low frequency temporal decays have a higher threshold and wider spread of values than high frequency temporal decays. In addition, lower SPL shows a higher threshold at lower frequencies.

8 ACKNOWLEDGEMENTS

The author would like to thank Dr. Aki Mäkivirta for his support and advice throughout the duration of the system development and the writing of this paper. In particular, he is credited with input on the headphone response correction and the chirp test signal algorithm. Prof. Matti Karjalainen has also offered overall guidance throughout this research. The listening test subjects are also thanked. Finally, Genelec Oy is thanked for the provision of technical facilities and support for this research.

9 REFERENCES

1. R. Bücklein, "The audibility of frequency response irregularities", *J. Audio Engin. Soc.* **29**, 126-131 (1981).
2. F. E. Toole and S. E. Olive, "The modification of timbre by resonances: Perception and measurement", *J. Audio Engin. Soc.* **36**, 122-142 (1988).
3. 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. Audio Engin. Soc.* **45**, 116-127 (1997).
4. B. J. C. Moore, "An introduction to the psychology of hearing, 5th edition", *Academic Press, London,* ISBN 0-12-505628-1.
5. B. M. Fazenda, M. R. Avis and W. J. Davies, "Difference limen for the Q-factor of room modes", *Proc. of 115th Audio Engin. Soc. Conf.*, preprint 5905 (2003).
6. M. M. Taylor and C. D. Creelman, "PEST: Efficient estimates on probability functions", *J. Acous. Soc. Am.* **41**, 782-787 (1969).
7. T. J. Cox and P. D'Antonio and M. R. Avis, "Room sizing and optimisation at low frequencies", *J. Audio Engin. Soc.* **52**, 640-651 (2004).
8. T. S. Welti and A. Devantier, "In-room low frequency optimisation", *Proc. of 115th Audio Engin. Soc. Conf.*, preprint 5942 (2003).
9. A. Goldberg and A. Mäkivirta, "Performance comparison of optimised equalisation techniques for in-situ loudspeaker measurements", *Reproduced Sound 21 Conf., Inst. of Acoustics, Oxford* (2005).

10. M. Karjalainen, E. Piirilä, A. Järvinen and J. Huopaniemi, “Comparison of loudspeaker equalisation methods based on DSP techniques”, *J. Audio Engin. Soc.* **47**, 14-31 (1999).
11. M. Karjalainen, T. Paatero, J. N. Mourjopoulos and P. D. Hatziantoniou, “About room response equalisation and dereverberation”, *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New York* (2005).
12. L. D. Fielder, “Analysis of traditional and reverberation-reducing methods for room equalisation”, *J. Audio Engin. Soc.* **51**, 3-26 (2003).
13. M. R. Avis, “IIR bi-quad controllers for low frequency acoustic resonance”, *Proc. of 111th Audio Engin. Soc. Conf.*, preprint 5474 (2001).
14. P. Antsalo, M. Karjalainen, A. Mäkivirta and V. Välimäki, “Comparison of modal equaliser design methods”, *Proc. of 114th Audio Engin. Soc. Conf.*, preprint 5844 (2003).
15. M. Karjalainen, P. A. A. Esquef, P. Antsalo, A. Mäkivirta and V. Välimäki, “Frequency-zooming ARMA modelling of resonant and reverberant systems”, *J. Audio Engin. Soc.* **50**, 1012-1029 (2002).
16. A. Mäkivirta, P. Antsalo, M. Karjalainen, and V. Välimäki, “Modal equalisation of loudspeaker-room responses at low frequencies”, *J. Audio Engin. Soc.* 51, 324-343, (2003).
17. M. R. Schroeder, “New method of measuring reverberation time”, *J. Acous. Soc. Am.* **37**, 409-412 (1965).
18. M. Karjalainen, P. Antsalo, A. Mäkivirta, T. Peltonen and V. Välimäki, “Estimation of modal decay parameters from noisy response measurements”, *J. Audio Engin. Soc.* **50**, 867-878 (2002).
19. M. C. Capp, J. R. Stuart and R. Wilson, “The loudspeaker-room interface – controlling excitation of room modes”, *Proc. of 23rd Audio Engin. Soc. Conf.* (2003).
20. M. Karjalainen, P. Antsalo, A. Mäkivirta and V. Välimäki, “Perception of temporal decay of low-frequency room modes”, *Proc. of 116th Audio Engin. Soc. Conf.* preprint 6083 (2004).
21. J. O. Smith, “http://www-ccrma.stanford.edu/~jos/filters/Pink_Noise_Filter.html#23473”, Dec. 2004.
22. H. Møller, D. Hammershøi, C. B. Jensen and M. J. Sørensen, “Transfer characteristics of headphones measured on human ears”, *J. Audio Engin. Soc.* **43**, 203-217 (1995).
23. G. Theile, “On the standardisation of the frequency response of high-quality studio headphones”, *J. Audio Engin. Soc.* **34**, 956-969 (1986).
24. H. Strasburger, “Invarience of the psychometric function for character recognition across the visual field”, *Percept. Psychophys.* **63**, 1356-1376 (2001).
25. H. Strasburger, “Converting between measures of slope of the psychometric function”, *Percept. Psychophys.* **63**, 1421-1455 (2001).
26. S. A. Klein, “Measuring, estimating and understanding the psychometric function: A commentary”, *Percept. Psychophys.* **63**, 1421-1455 (2001).
27. B. Treutwein and H. Strasburger, “Fitting the psychometric function”, *Percept. Psychophys.* **61**, 87-106 (1999).
28. The MathWorks, “MATLAB Optimisation Toolbox (v.3.0.1)”, *The MathWorks Inc., Natick,* (2004).
29. M. R. Leek, “Adaptive procedures in psychophysical research”, *Percept. Psychophys.* **63**, 1279-1292 (2001).
30. J. C. Nunnally and I. H. Bernstein, “Psychometric theory, 3rd edition”, *McGraw-Hill Inc., New York,* ISBN 0-07-047849-X.
31. C. Kaernbach, “Slope bias of psychometric functions derived from adaptive data”, *Percept. Psychophys.* **63**, 1389-1398 (2001).
32. B. R. Shelton, M. C. Picardi and D. M. Green, “Comparison of three adaptive psychophysical procedures”, *J. Acous. Soc. Am.* **71**, 1527-1533 (1982).
33. Wald, “Sequential analysis”, John Wiley and Sons Inc., New York, referenced in [6].
34. J. L. Hall, “Hybrid adaptive procedure for estimation of psychometric functions”, *J. Acous. Soc. Am.* **69**, 1763-1769 (1981).
35. M. M. Taylor, S. M. Forbes and C. D. Creelman, “PEST reduces bias in forced choice psychophysics”, *J. Acous. Soc. Am.* **74**, 1367-1374 (1983).
36. M. R. Leek, T. E. Hanna and L. Marshall, “Estimation of psychometric functions from adaptive tracking procedures”, *Percept. Psychophys.* **51**, 247-256 (1992).
37. The MathWorks, “MATLAB (v.7.0.1)”, *The MathWorks Inc., Natick,* (2004).