7 - 12 SEPTEMBER 2014, KRAKÓW
FORUM ACUSTICUM

Difference limen for reverberation time in auditoria

Andreas Tsolias and William J. Davies
Acoustics Research Centre, University of Salford, Salford, UK.

Summary
The smallest change in reverberation time that can be perceived in a concert hall has been measured. Five existing halls with average reverberation times varying between 0.89 and 2.3 s were simulated in detail with Bose room acoustic modelling software. The computer models were verified against acoustic measurements in the real halls. The models were then adapted to reduce the reverberation time in small steps and auralised with anechoic speech and music to generate a set of listening test stimuli. A Bose near-field rendition system was used to reproduce the simulated halls to ten listeners who made paired comparison judgements on the changing reverberation. The method of limits was adopted for the listening test, with a two-alternative forced-choice paradigm. An analysis of variance in the difference limen results was conducted for factors hall, subject, direction and motif. The bulk of the variance was due to the factor hall alone. When the difference limen are expressed as a fraction of the hall RT, then the variance is greatly reduced. This lends confidence to a single-figure difference limen of 3.2% +/- 0.6%. The result will be useful to those who design, simulate or auralise auditoria.

PACS no. 43.55, 43.66

1. Introduction

Reverberation time is the most important acoustic parameter in auditorium design. Hall designers know that it is difficult to predict reverberation time accurately. Uncertainties in the measurement of absorption coefficients and the failure of real rooms to meet the stringent assumption of a diffuse field are two of the several reasons why RT may not be predicted accurately. Nevertheless, the designer almost always has a target RT to try to meet. The question then arises of how close is good enough? Similar questions arise for the programmers and users of room acoustic modelling software: how near should the predicted RT be to that measured in the hall when it is built? We might also be interested in the significance of the variation in RT from one seat to another in the same hall. Are these small differences audible? One way of answering all three of these questions is to determine the smallest change in reverberation time that a listener can detect: the difference limen. If a measured or predicted RT gap is smaller than the difference limen, then we may assume that it will not be noticed. On the other hand, if a RT gap is, say, on the order of ten times the difference limen, then we should think of it as a large difference.

Several previous measurements have been made of the difference limen (or of the just-noticeable difference, or of the just-not-noticeable difference – what is actually measured depends on the test method employed). The results in the literature are summarised in Table I. The seminal work is due to Seraphim [1], who measured a difference limen using reverberated bands of noise. Seraphim found that the DL was a nearly constant fraction of the base RT. Although this was not measured with music or speech, Seraphim's result of 4% is the one most commonly used to provide advice to hall designers. ISO 3382 refers to Seraphim's work when stating a value of 5% for the difference limen [5].

Niaounakis and Davies [3] measured a value for the DL with very short RTs in a small studio control room. They found that, for short RTs, the DL did not vary much with the base RT and expressed the result as an absolute value of 0.042 s. Since then, several authors have measured a DL with music, speech and noise signals. The results vary from 3.3% to 26%. There may be two groups of results in Table I: small (3.3, 4 and 6%) and large (9.6, 10, 24.5 and 26%). Most of the

Table I. Previous results in the literature.

Author	*Experiment*	*Base RT (s)*	*Result*
Seraphim [1]	Noise, simple exponential decay	0.6 – 4 < 0.6s	4% 0.024 s
Karjalainen & Jarvelainen [2]	Noise, speech, simple exponential decay	0.5 – 2s	3.3 – 9.6%
Niaounakis & Davies [3]	Music, real room, 6 subjects	0.2 – 0.6	0.042 ± 0.015 s
Meng et al. [4]	Music, digital reverb, 30 subjects	1-4 s	26%
Frissen et al [6]	Noise, speech, simple exponential decay, 7-12 subjects	1.8 s	6%
Billon & Embrechts [7]	Noise, music, speech, one hall auralised, 15 subjects	1.89s	10%
Blevins et al. [8]	Oct band noise, one hall auralised, 4 subjects	1-3 s	24.5%

results in Table I use artificial reverberation, often a simple exponential decay. Billon and Embrechts [7] and Blevins et al. [8] argue that a more realistic test will come from using an auralisation of a computer model of a real hall. Both these authors did this, but for one particular hall only.

Given the range of results in Table I, it seems that the difference limen for RT is not a settled matter. The previous investigations have several merits, but they also have some shortcomings. The present paper seeks to improve understanding by adopting some of the best features of the previous studies – real hall auralisation and realistic signals. A significant extension in the results reported here is the auralising of several different halls.

2. Method

2.1. Auditoria

Five existing auditoria were chosen for the project. Summary data for the halls are given in Table II. An acoustic model of each hall was constructed in a commercially-available acoustic modelling package called Bose Modeller. The geometry and absorption coefficients of every significant surface in each hall was used as the basis of the model. Where possible, the reverberation time predicted by the model was compared with that measured in the hall. Figure 1 shows the shapes and reverberation times of the auditoria. Although the match between measured and predicted RT is not perfect, the models are adequate to produce an

Table II. Summary data for the five halls auralised

Abbrev.	*Name*	*Use*	*Shape*	*Volume (m^3)*	*Seats*	*RT (s)*	*RT range (s)*	*Auralised L_{Aeq} (dB)*
AD	Théâtre Alexandre Dumas, Saint-Germain, Paris, France	drama, opera, music	shoebox	9,898	670	1.35	1.00 – 1.35	71.5
KA	Kinnarps Arena, Jönköping, Sweden	multi-purpose	arena	81,688	4000	2.30	1.66 – 2.30	69.6
KO	Oslo Konserthus, Norway (small hall)	chamber music	shoebox	1,673	266	0.89	0.64 – 0.89	74.6
VT	Velinx Theater, Tongeren, Belgium	drama, music, dance	horseshoe	13,185	725	1.12	0.77 – 1.12	71.0
IM	Idrima Meizonos Ellinismou (theatre Antigone), Athens, Greece	drama, music	shoebox	9,538	895	1.00	0.74 – 1.00	73.1

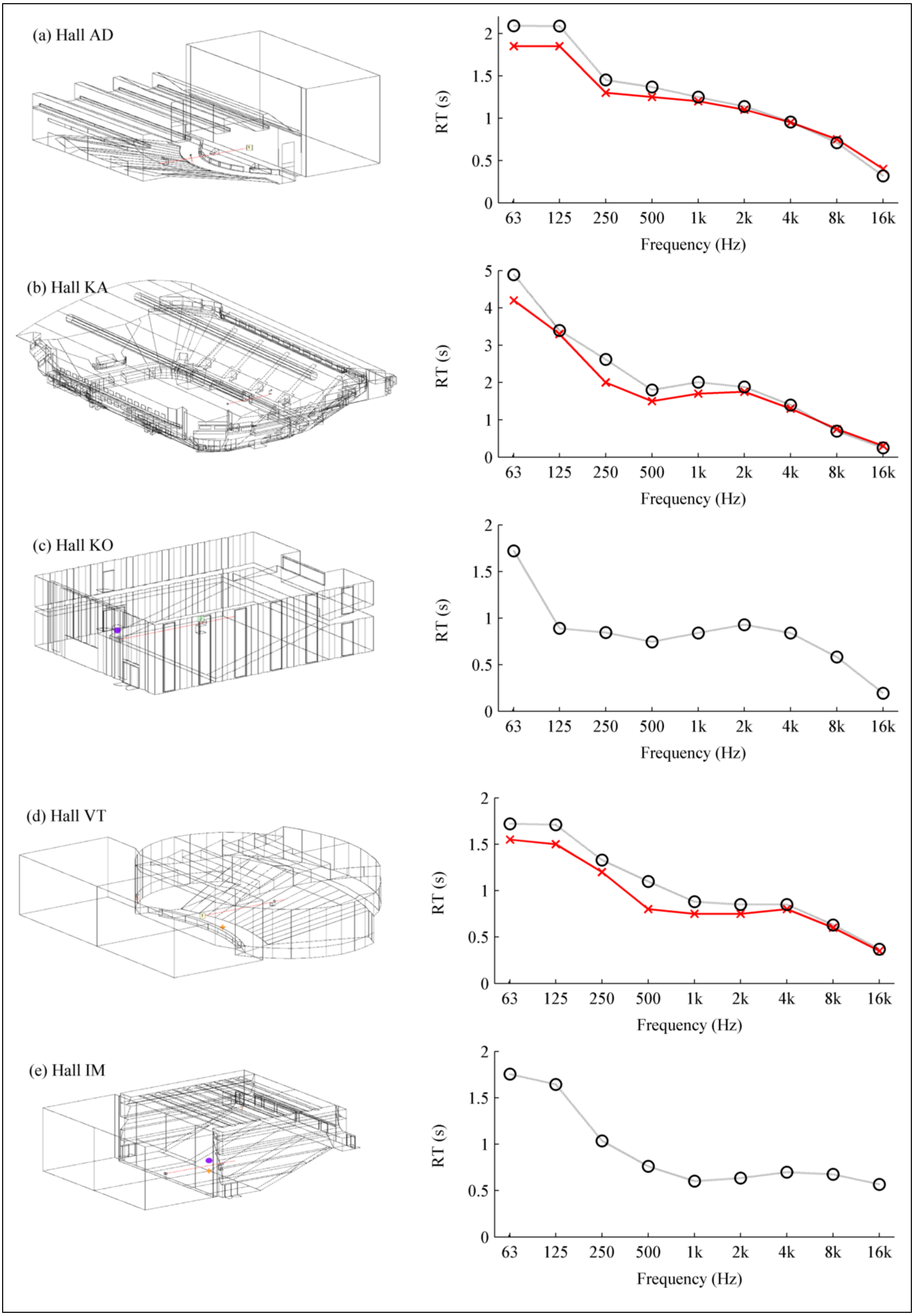

Figure 1. Isometric view of the five halls and their measured (x) and modelled (o) reverberation time.

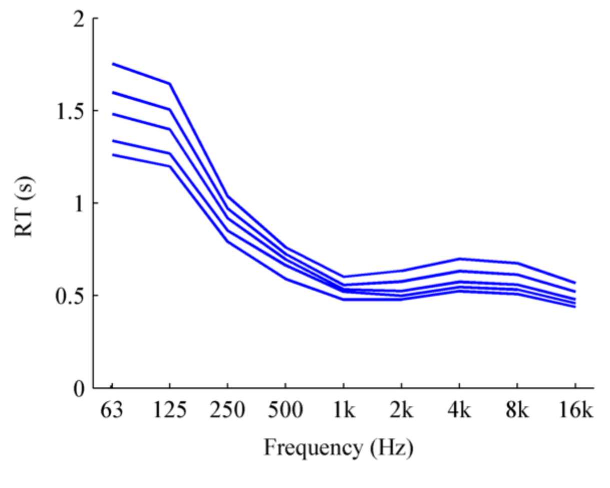

Figure 2. RT steps auralised in hall IM.

auralisation which sounds representative of the type of hall (large arena, small shoebox, etc.). The reverberation time of each modelled hall was changed incrementally by changing the absorption coefficient of several of the surfaces in the model. This produced auralisations which sounded natural and realistic, so that the range of reverberation heard is representative of halls which could be built. Figure 2 shows how RT changed in octave bands for hall IM, as an example. Table II shows the range of RTs used for each hall, averaged from 63 Hz to 16 kHz.

Varying an auralised model in this way produces realistic natural-sounding changes in reverberation. The experiment reported here was intended to examine the perception solely of changes in RT. Of course, the values of other parameters such as C_{80} and D_{50} will have changed along with the RT, because these parameters are all correlated in a real room. It would not be possible or desirable to change RT and nothing else. Because the changes to the absorption were evenly distributed in the room, the values of any other parameter should change in a realistic way. Moreover, the receiving position selected for the auralisation was well outside the room radius for each hall. Therefore, the single parameter which best describes the changes between the sound fields used here is the mean RT. In the statistical analysis below it is possible to examine whether mean RT is a good explanatory variable.

2.2. Stimuli

Two anechoic stimuli were used for auralisation. The speech stimulus was a 13-second segment of the loudspeaker designers Alan Shaw and Derek Hughes made in the BBC anechoic chamber at Kingswood Warren. They are having a conversation about speech reproduction. The music stimulus was a 16-second section downloaded[2] from a recording of Bruckner's symphony no. 8, movement II, bars 1-61. This recording was made for auralisation test purposes [9]. The anechoic stimuli were equalised to give the same L_{Aeq} at the input to each model hall. However, the source-receiver distance and strength of the halls varied, so that the level at the receiver varied. The listener level for each hall is shown in Table II. This variation is a consequence of using a range of hall sizes and it was decided that the fairly small range of levels was part of the realism of the experiment. Therefore, the final listener levels were not equalised across the halls.

2.3. Reproduction

The reproduction system was unusual. A commercial Bose Auditioner III system was used. This employs a near-field loudspeaker system of two tweeters and two woofers positioned 20-25 cm away from the listener's ears [10]. The listener rests their chin on a rest, so a reasonably consistent head location is obtained. A sound level meter is integrated into the base of the device so that the sound level can be measured during playback. The playback system was sited on a desk in an untreated office. The office had a volume of 162 m^3, a mean RT of 0.57 s and a background noise level L_{Aeq} of 41 dB. Because this environment is not a controlled listening or anechoic room, there is a possibility that the reverberation or noise in the playback room could have interfered with what the listeners heard. However, the listener's ears are only 20 cm from the playback loudspeakers and on axis, so the signal at the ear will be dominated by the auralisation rather than the playback room reverberation. The reproduction system is perhaps closer to headphones than to normal stereo loudspeakers. Even with conservative estimates, the playback room reverberation will be at least 20 dB below the auralisation reverberation. Combined with a background noise level of 41dB, it seems likely that listener judgements will be determined directly by the auralised signals, as intended.

2.4. Listeners

Ten participants took part. Each had a normal audiogram. Four could be described as expert listeners, because they work as acoustic

[2] http://auralization.tkk.fi/node/13

Table III.Sources of variance in the absolute difference limen: Four-way anova with second-order interaction.

Source	*d.f.*	*F*	*p*
Dir	1	19.28	0.0063*
Hall	4	103.74	0*
Motif	1	4.44	0.3101
Subject	7	1.8	0.5329
dir*hall	4	5.1	0.0026*
dir*motif	1	1.55	0.2222
dir*subject	7	0.89	0.5286
hall*motif	4	1.52	0.2195
hall*subject	22	1.72	0.0778
motif*subject	6	0.43	0.8552
Error	33		
Total	91		

consultants. The other six listeners should be classed as naïve. Ages ranged from 18 to 53 with a mean of 41. Seven were male and three female.

2.5. Test method

The method of limits was used, with a 2AFC paradigm. This is an efficient method for finding both just-noticeable and just-not-noticeable differences. Listeners were presented with a sequence of two auralised signals and asked to choose different or not different. The first pair presented a large difference in RT and subsequent pairs reduced the difference until the point of equality was reached. The jnnd was recorded as the RT halfway between the different/same responses. Starting again from the point of equality, the pairs increased in RT difference until the listener reported hearing a difference. The jnd was recorded as the RT halfway between the same/different responses. This method therefore produces a 50% limen: the RT at which there is a 50% chance of a difference being heard. Errors of habituation and anticipation can be expected to influence the judgements, so the jnd and jnnd are usually different. The normal procedure in the method of limits is to estimate the true difference limen from the average of the jnd and jnnd.

The order of the halls was randomised, but the motif was not: speech was always presented before music for each hall and subject. In an informal pilot test, it was easier to be confident

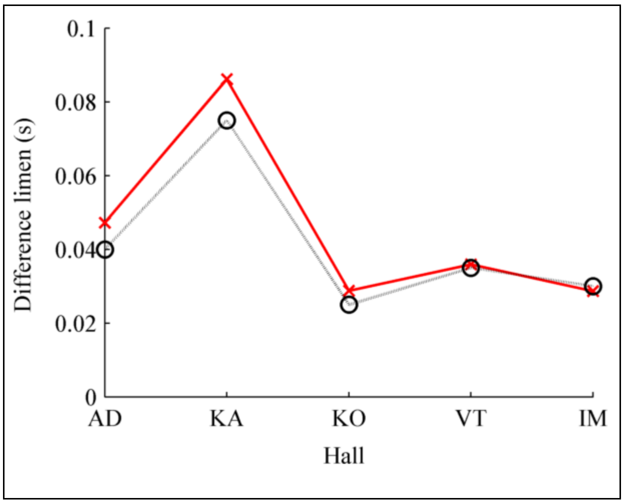

Figure 3. Mean just-noticeable (o) and just-not-noticeable (x) difference in RT for each hall.

when judging speech, as opposed to music. Fewer errors resulted if participants got used to listening to the hall with speech before tackling it with music. It is possible, therefore, that any order effects could confound any effect of motif.

3. Results and discussion

The subjects found the experiment quite difficult, and it was more difficult or music than for speech. This may be because the Bruckner piece used is complex with a broad spectral and dynamic range. The difficulties are reflected in the dataset, since there are several missing values where subjects reported hearing a difference at the point of equality. Two subjects made large numbers of such errors and have been excluded from the analysis. A four-way analysis of variance was conducted on the reported difference limen in seconds. At this stage, none of the subjective data was averaged. The four factors were: direction (jnd, jnnd), hall (AD, KA, KO, VT, IM), motif (speech, music) and subject (8 subjects). Subject was treated as a random factor and the other three factors as fixed. Notice that hall is used as a fixed factor instead of RT as a linear factor. This allows for the possibility that the effect of the hall is due to something more than, or different to, their different mean RT. The first analysis fitted a model with up to third-order interactions (e.g. hall*motif*subject). None of the third-order interactions were significant ($p>0.75$ for all).

The anova was then re-run to fit a model with up to second-order interactions. (This treats the third-order interactions as part of the random error.) The results are shown in Table III. The only second-order interaction which is significant

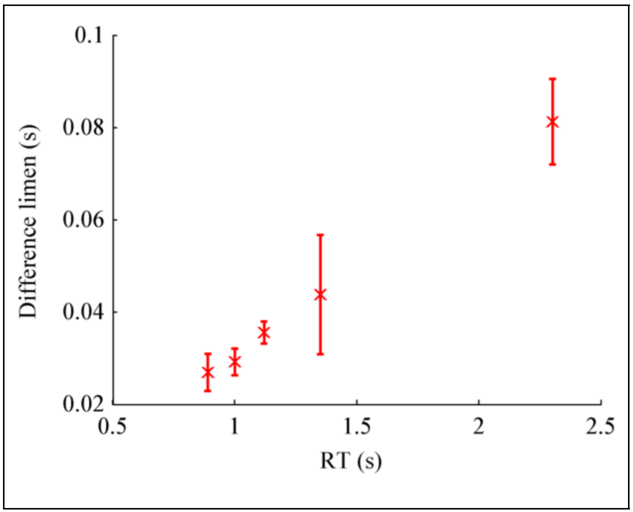

Figure 4. Absolute difference limen as a function of hall mean RT.

(p<0.05) is dir*hall. This interaction is examined by plotting mean DL against hall for the two directions in Fig. 3. The lines do cross, confirming the interaction. This means that caution should be exercised in interpreting the main effect of the factors direction and hall.

Table III shows that motif and subject are not significant, and that direction and hall are significant (p<0.05). It seems clear from Fig. 3 that the jnnd and jnd are close and should be averaged to produce a sensible estimate of the difference limen. This leaves hall as the only remaining significant source of variance in the data. The five halls have different mean RTs, so the mean DL for each hall can be plotted against its RT in Fig. 4. The linearity of this plot strongly suggests that the variance caused by the factor hall

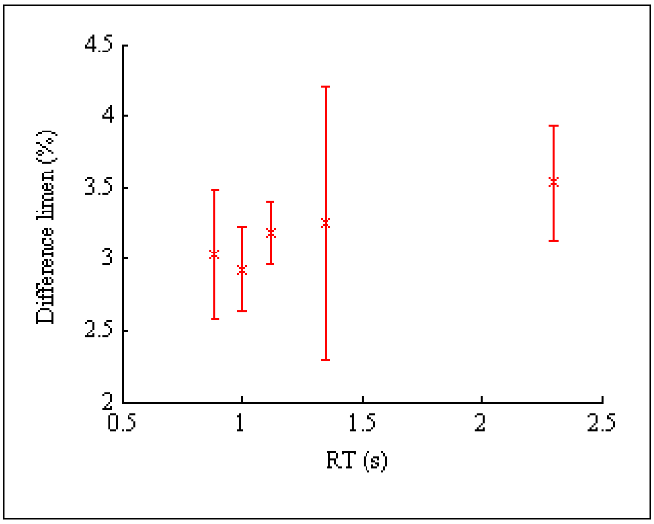

Figure 5. Percentage difference limen as a function of hall mean RT.

is due solely to the halls having different reverberation times. Figure 4 also answers the question raised in section 2.1 above. It seems very likely that the changes in each auralised model were perceived as changes in RT and so this aspect of the method is justified.

It therefore seems sensible to express the difference limen as a percentage of the hall RT, as some previous authors have. The original response data were therefore transformed to percentages and the analysis of variance repeated. In the new anova, the factor hall was replaced by RT, treated as a continuous factor. The new anova for percentage DL showed that some variance still remained in the data, but that the only significant factor in the variance was RT, with p=0.03. This is examined by plotting the mean percentage DL against RT in Fig. 5. While there is a slight positive gradient, the variation of percentage DL with RT is small and does not seem to justify any further transformation of the data.

Taken together, the results support expressing the DL as a single mean percentage of the hall RT. This results in a final mean difference limen of 3.2% +/- 0.6%. This result fits quite well with the 'small' values in Table I: 4% (Seraphim [1]), 6% (Niaounakis and Davies [3]), 6% (Frissen et al. [6]) and the lower range of Karjalainen & Jarvelainen [2] (3.3%). The 'large' group in the literature are at odds with these data, with values of 10% (Billon and Embrechts [7]), 24.5% (Blevins et al. [8]) and 26% (Meng et al. [4]). The finding that motif is not significant agrees with the results of Niaounakis and Davies [3] and those of Frissen et al. [6], though Billon and Embrechts [7] found a small effect.

4. Conclusions

The difference limen for reverberation time has been measured using realistic auralisations of five existing halls. The choice of music or speech did not have a significant effect on the result. A strong linear relationship was found between absolute difference limen and mean reverberation time over the five halls. This suggests that some confidence can be given to an overall mean difference limen of 3.2% +/- 0.6%.

This result is useful because it updates the original work of Seraphim (based on simple reverberations of noise bands), which itself is still used as the basis of current advice to acoustic designers and researchers. The new value for the difference limen will be useful to establish a yardstick for the necessary accuracy of reverberation time measurements and predictions. The difference limen can be used as a target for the error margin in a hall design, as a comparator for RT variance across seat positions, as a guide to the required accuracy in room acoustic simulation and auralisation, and as a measurement of the capability of this aspect of the human auditory system.

References

[1] H. Seraphim: Untersuchungen über die unterschiedsschwelle exponentiellen abklingens von raushbandimpulsen, Acustica 8 (1958) 280-284.

[2] M. Karjalainen, H. Jarvelainen: More about this reverberation science: Perceptually good late reverberation, in: Audio Engineering Society Convention 111, New York, U.S.A, 2001.

[3] T. I. Niaounakis, W. J. Davies: Perception of reverberation time in small listening rooms, Journal of the Audio Engineering Society 50(5) (2002) 343-350.

[4] Z. Meng, F. Zhao, M. He: The just noticeable difference of noise length and reverberation perception, in: International Symposium on Communications and Information Technologies (ISCIT'06), 2006.

[5] ISO3382-1: Acoustics – Measurement of Room Acoustic Parameters – Part 1: Performance Spaces. 2009.

[6] I. Frissen, B. F. G. Katz, C. Guastavino: Effect of Sound Source Stimuli on the Perception of Reverberation in Large Volumes, in: S. Ystad, et al., (Eds.),Auditory Display, 2010 pp. 358-376.

[7] A. Billon, J.-J. Embrechts: Discrimination thresholds of the reverberation in large volumes by naive listeners, in: Acoustics 2012, Nantes, France, 2012.

[8] M. G. Blevins, A. T. Buck, Z. Peng, L. M. Wang: Quantifying the just noticeable difference of reverberation time with band-limited noise centered around 1000 Hz using a transformed up-down adaptive method, in: International Symposium on Room Acoustics, Toronto, Canada, 2013.

[9] T. Lokki, J. Pätynen: Applying anechoic recordings in auralization, in: Proc. of the EAA Symposium on Auralization, Espoo, Finland, 2009.

[10] M. Jorgensen, C. B. Ickler, K. D. Jacob: Using Subject-Based Testing to Evaluate the Accuracy of an Audible Simulation System, in Audio Engineering Society Convention 95, 1993.