PERCEPTION OF REVERBERATION IN LARGE SINGLE AND COUPLED VOLUMES

Ilja Frissen[1], Brain F.G. Katz[2], Catherine Guastavino[1]

[1]Center for Interdisciplinary Research in Music, Media and Technology,
McGill University, Montreal, Canada
[2]LIMSI-CNRS, Orsay, France

ilja.frissen@mcgill.ca
brian.katz@limsi.fr
catherine.guastavino@mcgill.ca

ABSTRACT

The aim of the presented research is to quantify how sensitive the human ear is to subtle changes in reverberation. We quantified the discrimination thresholds for reverberations that are representative for large rooms such as concert halls (reverberation times around 1.8 s). For exponential decays, simulating an ideal simple room, thresholds are around 6% (Experiment 1). We found no difference in thresholds between a short noise burst and a male voice spoken word, suggesting that discrimination is not dependent on the type, or spectral content, of the sound source (Experiment 2). In a final experiment we matched coupled room, non-exponential decay stimuli to exponential ones, and vice versa, in an attempt to quantify the complex former in terms of the simpler latter.

1. INTRODUCTION

Reverberation is the collection of reflected sounds from the surfaces in an enclosed volume. The direct sound is followed by distinct reflected sounds and then a collection of diffuse reflections which blend and overlap into what is called “reverberation”. Reverberation is a crucial acoustical parameter characterizing the sound quality of an auditorium. In our daily lives we are continually confronted with echoic and reverberant environments. Not only do we live in them, we also evolved in them; the ancestral cave comes to mind.

Despite their being ubiquitous there is surprisingly little research on the perceptual processing of reverberant stimuli [see 1]. In those studies that have been done, exponentially decaying reverberation profiles are used to simulate what would happen in a single room. The primary parameter for characterisation of importance is the reverberation time (RT) which is defined as the time required for reflections of a direct sound to decay by 60 dB below the level of the direct sound. Seraphim [2], for instance, determined the discrimination thresholds for a narrowband noise burst with RTs ranging from 170 ms up to 10 s, and found the best discrimination (3-5%) between RTs of 0.8 and ~4 s. He also stated that this performance neither depended on the frequency of the stimulus nor on its length. Niaounakis and Davies [3] investigated perceptual thresholds in relatively small rooms (RT < 0.6 s) using very long (i.e., 21 s) musical excerpts. Interestingly, despite major differences in stimulus type, they found thresholds similar to [2], around 6%. This allows the speculation that the discrimination of reverberation is not strongly dependent on the semantics or even the spectral content of the sound source.

In our everyday life the listening environment is often not a simple single room but often a number of connected rooms. Take, for instance, an office with its door open into a hallway, or a concert hall, with the orchestra pit as the primary and the audience area as the secondary room. [4] and [5] showed that intensity does not decay linearly over time in actual concert halls but rather exhibits a double (or multiple) slope decay. Some halls are designed specifically to behave in this manner, as shown in such examples as [6-7]. Non-exponential decays occur when an adjacent reverberant volume(s) is connected to a less reverberant space containing both the acoustic excitation source and the acoustic receiver. Under certain conditions, the resulting energy exchange between different rooms coupled via doorways or apertures can result in significant non-linear exponential decays. Considering the simple case of a double slope decay (two coupled spaces), the *non-exponential* decay can be simulated by adding the two simple exponential decay functions corresponding to the individual rooms, with their individual RTs [4]. In addition, the point at which the two slopes intersect (at a certain time after onset of the direct sound and at a certain level of attenuation) becomes an important parameter. Recently, these more complex environments have also started to receive scientific attention [5, 8-9].

A thorough understanding of human sensitivity to reverberant stimuli is not only of theoretical value but has large practical potential as well. It will instruct the architecture of rooms specifically intended for listening, such as concert halls. But also the acoustic display of information would benefit. A random example of a potential application is the use of simulated rooms of different sizes to cluster and/or categorize like information.

This work presents the first steps in gaining this understanding. It gives the results of three separate experiments investigating human perception (discrimination) of reverberation in *large* rooms. Accordingly we restricted the investigation to overall reverberation times to RTs on the order of 1.8 s, which is representative for many concert halls [4]. In the first experiment we determine the discrimination sensitivity of synthetic single-room, or *exponentially decaying reverberation*, profiles. In the second experiment we examine the observation that discrimination thresholds are apparently not dependent on the type of stimulus used. Finally, we introduce synthetic coupled-room, or *non-exponential decay reverberation*, profiles, and try to find a perceptual match of exponential decay to a non-exponential decay profiles, and vice versa.

2. EXPERIMENT 1

The first experiment was a replication of that part of Seraphim's results [2] that is of relevance here. Thus we only tested an RT in the region of 1.8 s, which in Seraphim's study produced a JND of about 4%.

2.1. Methods

2.1.1. Participants

Eight participants, including two of the authors, completed the experiment. All participants were tested for normal hearing using a standard audiometric test (over the range of 250 - 8000 Hz).

2.1.2. Stimuli

A series of nine IRs, with an exponential decay, were synthesized using custom written software developed by one of the authors in Matlab (the MathWorks). The IR is the result of the application of a simple exponential decay to a normally distributed random number sequence. The RT was varied from 1.48 to 2.21 s in equal steps of ~11 ms. In order to obtain the experimental stimulus, the IRs were convolved with a 170 ms white noise burst. Sounds were presented over headphones (AKG K-271 Studio), played from a MacPro connected to an audio interface (Motu, mkII 828) at a sample rate of 44.1 kHz.

2.1.3. Procedure

The experiment employed a standard method of constant stimuli paradigm, with a 2-interval, forced-choice task. The task for the participant was to judge "which one had the most reverberation". The stimulus with an RT of 1.8 s served as the reference. Each of the nine comparisons was tested 12 times for a total of 106 trials. To reduce any inadvertent effects of response biases, the order of presentation of the reference and comparison was randomized such that in half of the trials the reference was presented first.

Presentation of the stimuli was controlled through a simple graphical user interface (GUI) developed and run in Matlab. The GUI featured only three buttons. One was a large "play" button that would play the stimulus pair with a random pause (0.5 – 2 s) in between. The participants were free to listen to the stimulus pair as many times as desired (although this option was used only rarely). To enter their response, they clicked one of two buttons corresponding to the "first" and "second" sound in the stimulus pair.

2.2. Results and Discussion

The proportion of trials in which each comparison was perceived to have more reverberation as the standard was calculated. To obtain psychometric functions, the data were fitted with cumulative Gaussians free to vary in position and slope using the software package *psignifit* (see http://bootstrap-software.org/psignifit/; [10]). The discrimination threshold (or, just-noticeable-difference, JND) was determined from the slope of the psychometric function. It was defined, as per convention, as the difference between the RTs that correspond to the 75% and 50% points of the cumulative Gaussian. Thus, the steeper the psychometric function, the more sensitive the corresponding discrimination, and therefore the smaller the JND.

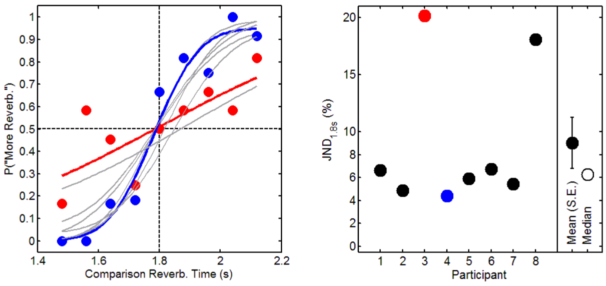

Figure 1. *Experiment 1. Left panel. Individual psychometric functions. Highlighted in color the curves for best (blue) and worst (red) performance. Right panel. Individual thresholds (JNDs) with same color coding and their averages.*

The results are summarized in Figure 1, with the left panel showing the individual psychometric functions. What can be seen is that six out of the eight participants showed very similar performance. Participants 3 and 8, on the other hand, found the task much more demanding as evidenced by their relatively flat psychometric functions. To illustrate the difference, the best and worst individual performances are highlighted in blue and red, respectively. The individual JNDs, and their average(s), are shown in the right panel. The mean JND was around 9%. However, because this value is heavily skewed by the two 'outliers', we chose to utilize the median, which is approximately 6%, very similar to the results presented in [2] and [3].

3. EXPERIMENT 2

In the introduction we noted that similar discrimination thresholds were found by [2] who used a simple narrowband noise and [3] who used complex musical excerpts. This similarity suggests that the nature of the sound source is not that important. The large methodological differences, of course, do not warrant a definite conclusion. We therefore tested the hypothesis directly in a single experiment.

3.1. Methods

Except for the following details all was the same as in Experiment 1.

3.1.1. Participants

Four of the participants in Experiment 1, including the first author, completed the experiment.

3.1.2. Stimuli

Sound sources were the 170 ms noise from Experiment 1 and a 600 ms recording of the French word "poussez", which was extracted from a recording made in a dry room of a French male speaker. The voice stimuli are the noise stimuli we calibrated in level to have the same root-mean-square value. As in Experiment 1, the IRs had a linear exponential decay.

3.1.3. Procedure

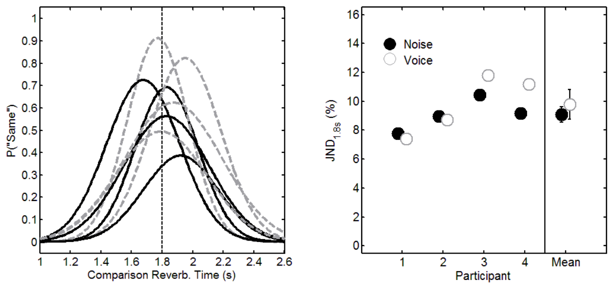

Figure 2. *Experiment 2. Left panel. All individual psychometric functions for noise (black lines) and voices (grey dotted lines) Right panel. Individual thresholds (JNDs) and their averages.*

The reference RT was 1.8 s and the tested RTs for comparison ranged from 1.48 to 2.12 s in seven equal steps of 80 ms. Each comparison for both sounds was presented nine times, for a total of 126 completely randomized trials. This time the participants were to judge whether the pair of stimuli were the same or different from each other.

3.2. Results and Discussion

From the raw data we calculated the proportion of "same" responses for each comparison and fitted a Gaussian function. The standard deviation of the Gaussian corresponds to the JND. The obtained curves are shown in the left panel of Figure 2.The right panel shows the individual JNDs and their mean.

There was no statistically significant difference between the JND for the Noise (9.1%) and the Voice (9.8%) stimuli, $t(3) = 1.17$, $p > 0.32$.

Thus, it seems that the type of stimulus does not affect the discriminability of reverberation. Obviously the generality of this statement is constrained, and further comparison using a broader range of stimuli is needed to find its limits.

The thresholds are noticeably higher than in Experiment 1. This difference turned out to be significant by a paired t-test, $t(3) = 9.04$, $p < 0.01$. The difference is most likely attributed to the nature of the task, whereas in Experiment 1 people were forced to directly compare the amount of reverberation, in Experiment 2, they 'merely' had to decide whether the stimuli were the same or different. It seems that they used a less conservative criterion in performing the latter task.

4. EXPERIMENT 3

Trying to find discrimination thresholds of non-exponential decays poses a problem. As mentioned in the introduction, coupled room reverberation can be simulated by combining the exponential decays from the two individual rooms (see Figure 3). To define the total decay, we need several parameters: the RT of the primary (RT early) room from which the sound source originates, the RT of the secondary room (RT late), and their respective level difference. Their relation determines the intersection of the two, which is specified in time and level, where the change in slope of the decay occurs. The problem for the normal psychophysical paradigm lies in the interdependence of the parameters in real physical rooms. Ideally one would want to be able to manipulate each parameter independently (through simple synthesis simulation). This "ideal" method then produces decays which are not necessarily realisable in an actual given geometry (or geometrical simulation) where it is not possible to change one without affecting at least one of the others.

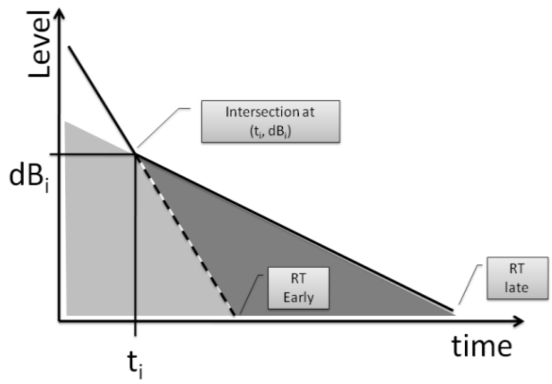

Figure 3. *Illustration of a non-exponential decay profile as typical for coupled rooms (see also text).*

We therefore decided not to determine a discrimination threshold as such. Instead we created two scenarios. In the first, we varied the intersection point along the early RT slope while keeping the total time constant at 1.8 s. This was achieved by varying t_i, and consequently the starting level, of the second decay). This way we wished to determine at what (intersection) point a non-exponential decay *starts sounding different* from an exponential one. In the second scenario we tried to characterize at what point an exponential decay is comparable to a non-exponential decay (cf. [8]). These two scenarios require somewhat different dependent measures. For the first we look at the point where a non-exponential decay profile is perceived to be different for 75% of the cases. For the second, we are looking at the point of subjective equivalence (PSE). We express these measures in terms of RT_{late} for scenario 1 and RT of the exponential decay in scenario 2.

Because of the task differences we found between Experiments 1 and 2 we ran Experiment 3 using both tasks, with half the participants performing the 2IFC task and the other half the same/different task.

4.1. Methods

4.1.1. Participants

Sixteen people, including the first author, completed the experiment.

4.1.2. Stimuli

The scenarios are illustrated in schematic form in Figure 4. In the first scenario the reference was an exponential decay with an RT of 1.8 s and the comparisons were non-exponential IRs. The parameter that was varied was the TR_{late}, by moving the along the slope of the early decay part (t_i: between 70 and 400 ms, with corresponding dB_i values: -4.2 and -24 dB). These values correspond to RT_{late}'s between 1.93 to 2.73 s.

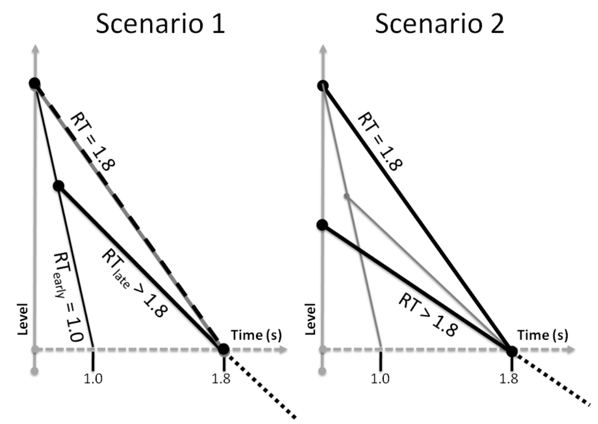

Figure 4. *Illustration of the stimuli used in the two scenarios in Experiment 3. The grey lines indicate the standard, and fat black lines show examples of comparisons. In scenario 1 the standard and one of the comparisons were identical (black/grey dotted line). The position of the intersection point was varied, which created variable RT_{Late}'s. For scenario 2 there was one double slope standard and a range of single slope comparisons with varying RTs and starting levels. The simulated Impulse Responses were cut last 1.8s.*

In the second scenario we had one, coupled room, non-exponential decay, as reference, with an early and late RT of 1.0 and 1.8 s, respectively and an intersection point at 150 ms (with a corresponding attenuation of ~-7.5 dB). As comparisons we used a range of exponential decays with starting levels varying between 0 and -15 dB. The corresponding RTs were chosen such that all the slopes crossed -60 dB at the same point as the reference. Thus the comparison RTs ranged from 1.8 to 2.4 s.

IRs were cut to last 1.8 seconds and convolved with a cropped version (to 350 ms) of the voice stimulus in Experiment 2, which turned out to sound like */te/*.

4.1.3. Procedure

The general procedure was very similar to Experiments 1 and 2. Each of the nine comparisons in both scenarios was presented eight times for a total of 112 completely randomized trials. Two versions of the same experiment were run, with eight participants assigned to each. The difference lay in the psychophysical task. In the first, the instruction to the participants was to decide "which sound had the most reverb". In the second, participants were asked if the two sounds the "same or different amounts of reverberation".

4.2. Results and discussion

Thresholds were obtained in a similar fashion as in the previous experiments. A summary of the results are shown in Table 1. For scenario 1 we find that on average a non-exponential decay with a late reverberation time of ~2.9 s (intersection point at 385 ms) is perceived to be different, 75% of the time, from the RT of 1.8 s in the reference..

For scenario 2 we see that the non-exponential decay profile can be perceptually matched with an exponential decay with an RT of ~1.9 s.

Scen.	2IFC	Same/ Different	Mean
1	2.53 (0.07)	3.38 (0.67)	2.92 (0.32)
2	1.80 (0.04)	1.90 (0.15)	1.85 (0.08)

Table 1. *Mean RT values corresponding to the threshold (Scenario 1) and PSE (Scenario 2) plus standard error of the mean, per task. The last column shows the average across tasks.*

As it turned out, the tasks were more difficult to perform for a number of participants than was anticipated. In some cases it was not possible to obtain reliable fits and the dependent measures were consequently considered a missing value and therefore did not enter into the calculation of the group means. These cases were due to the fact that the participant was unable to discriminate the stimuli in a systematic fashion. One explanation for this is that the stimuli were too dissimilar that it was hard for the participants to come up with a consistent criterion by which to make their judgement.

5. GENERAL DISCUSSION

This study presents an investigation of the human perception of reverberation in large rooms, through a series of experiments. It was found that RT discrimination is relatively sensitive with thresholds around 6% (Experiment 1), which is in agreement with earlier findings [2-3].

The experimental confirmation that stimulus type does not appear to affect its perceptual discrimination processing (Experiment 2) allows an interesting insight. It suggests that the perceptual processing of reverberant inputs occurs before the semantics of the stimulus are appreciated.

The results of Experiment 3 in which we tried to find exponential equivalents of non-exponential profiles and vice versa are less clear. Although this partial ambiguity in the individual results was already anticipated in the section's introduction, it was desired to use quantitative methods in order to characterize coupled room reverberation perception.

It seems that one of the obstacles is to provide the participant with a discrimination task that is sensible. It was thought that a "same/different" task would do just that, since it does not require the explicit comparison of the two stimuli on a particular perceptual dimension (e.g., decay time). Surprisingly, this task produced higher thresholds (as in Experiment 2) and was still too ambiguous for some participants in Experiment 3.

The more general issue is the multidimensional nature of the coupled room stimuli. Thus, reasonably, future efforts will require a step away from conventional psychometric methodology and towards methods that will allow us to deal with such complex stimuli where comparisons for physically realisable non-exponential decays are studied. The perceptual component will also be extended directly to musical applications and address perceptual preferences, rather than simply perceptual discrimination tasks.

6. CONCLUSION

The work reported here is only the first step in a trying to better understand how the perceptual system deals with reverberant environments. Its study is of theoretical importance because reverberant environments are ubiquitous, and it is well known that they affect (both beneficially and adversely depending on the situation) sensory processes. A better understanding then might provide some insight into the ways the brain has adapted to cope (or not) with reverberant stimuli.

There is also some practical importance in that understanding how the brain is able to deal with reverberant environments could create novel ways of auditory information display. Particularly here the apparent failure of conventional psychophysics puts some urge to finding new ways of quantifying the perception of (coupled) room reverberation.

7. REFERENCES

[1] B.F.G. Katz, "International round robin on room acoustical impulse response analysis software 2004". *ARLO,* vol. 5, no. 4, pp. 158-164, 2004.

[2] H. P. Seraphim, "Untersuchungen über die unterschiedsschwelle exponentiellen abklingens von raushbandimpulsen", *Acustica*, vol. 8, pp. 280-284, 1958.

[3] T.I. Nianounaki, and W.J. Davies, *J. Audio Eng. Soc.*, vol. 50, pp. 343-350, 2002.

[4] E. Kahle, *"Validation d'un modèle objectif de la perception de la qualité acoustique dans une ensemble de salles de concerts et d'opéras"*. Unpublished Ph.D. dissertation, Université du Maine, Le Mans, 1995.

[5] D.T. Bradley and L.M. Wang, "The effects of simple coupled volume geometry on the objective and subjective results from exponential decay", *J. Acoust. Soc. Am.*, vol. 118, no. 3, 2005.

[6] B.F.G. Katz and E. Kahle, "Design of the new Opera House of the Suzhou Science & Arts Cultural Center." *Proceedings of the 9th Western Pacific Acoustics Conference, WESPAC IX,* Seoul, Korea, 26-28 June, 2006.

[7] E. Kahle, R. Johnson, B.F.G. Katz, "The new konzertsaal of the KKL Center, Lucerne, Switzerland. II Preliminary acoustical measurements." *Acta Acustica* vol. 85, pp. S2, 1999.

[8] D. Picard, *"Audibility of non-exponential reverberation decays"*, Unpublished thesis, Rensselaer Polytechnical Institute, Troy, NY, 2003.

[9] K.S. Sum and J. Pan, "Subjective evaluation of reverberations times of sound fields with non-exponential decays", *Acta Acust United Ac*., vol. 92, pp. 583-592, 2006.

[10] F.A. Wichmann and N.J. Hill, "The psychometric function: I. Fitting, sampling, and goodness of fit", *Percept. Psychophys*, vol. 63, no. 8, 2001.

ACKNOWLEDGEMENTS

We thank Aaron Rosenblum for helping with the data collection. This work was supported by FQRNT grant (#113581) and CFI grant (#11367) to C. Guastavino. Testing took place at the McGill Multimodal Interaction Laboratory and at the Centre for Interdisciplinary Research on Music Media and Technology.