



# Audio Engineering Society Convention Paper 6084

Presented at the 116th Convention  
2004 May 8–11 Berlin, Germany

*This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42<sup>nd</sup> Street, New York, New York 10165-2520, USA; also see [www.aes.org](http://www.aes.org). All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.*

---

## Psychoacoustic Cues in Room Size Perception

Sharaf Hameed<sup>1</sup>, Jyri Pakarinen<sup>1</sup>, Kari Valde<sup>2</sup>, and Ville Pulkki<sup>1</sup>

<sup>1</sup>Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, P.O. Box 3000, FIN-02015 HUT, Espoo, Finland.

<sup>2</sup>Helsinki University of Technology, P.O. Box 1000, FIN-02015 HUT, Espoo, Finland.

Correspondence should be addressed to Sharaf Hameed ([sharaf.hameed@hut.fi](mailto:sharaf.hameed@hut.fi))

### ABSTRACT

The ability of human listeners to estimate the size of a room from the acoustical response of that room is an interesting and not yet thoroughly examined phenomenon. This study uses simulated multi-channel room impulse responses convolved with speech signals as stimuli in listening tests to explore the perception of room size. The synthetic room impulse responses contained two adjustable parameters, and our goal was to study how these parameters affect the perceived size of this virtual room. Listening tests were conducted to test the effect of reverberation time and the direct to reverberant energy ratio (D/R ratio). Sound samples with different parameter settings were presented as stimuli in a paired comparison test procedure. The results reveal that reverberation time is unequivocally the most important parameter. It appears that D/R ratio is not used in room size perception.

### 1. INTRODUCTION

An auditory stimulus in different spaces yields differ-

ent auditory responses, depending on the geometry of the space, and on positions of the sound source

receiver. The direct sound from the sound source arrives at the listeners ears first followed by reflections and reverberation. The impulse response is measured in the listening position with an omnidirectional microphone. An omnidirectional microphone has (almost) equal sensitivity in all directions. The response gives us information about arrival times of direct sound and reflections and about the magnitude of reverberation. A typical room impulse response is illustrated in Figure 1.

When a continuous sound is presented to a listener, direct sound arrives first. Before reflections arrive, the direction of sound can be decoded, since binaural directional cues carry the information about it. When reflected sound arrives from all directions surrounding the listener, it is summed to the direct sound, forming cues that may deviate significantly from cues of the direct sound. The auditory mechanisms binaurally decode the cues present in ear canal signals to form an auditory percept of the sound source at some point in the listening space.

Since humans are able to perceive the size of the room from its auditory response [1], it is clear that some features of the impulse response are responsible and act as cues in the estimation of the size of the room. If the distance between the sound source and the perceiver and their orientations is fixed, it is reasonable that the direct sound does not depend on the surrounding environment, and does not contribute to the perceived size of the room. This turns our interest towards the early reflections and the reverberant part.

Previous studies [3] have shown that listeners are somewhat sensitive to the early and late reverberant part of sound and are able to extract some of the information from the gross characteristics of the reverberation, but they are not sensitive to the exact pattern or timing of early reflections. It was shown that they are also unable to extract information about the position of the speaker or listener in the room from the reverberation. The early reflections, or more specifically, the time delays between them and the direct sound clearly correlate with the size of the room; in a small room the reflections arrive sooner to the listener than in large rooms. The amplitudes of the early reflections also depend on the size of the room, as the sound level is inversely proportional to the distance traveled. How-

ever, since it was established in [3] that humans are incapable of discriminating between the exact amplitude, timing and direction of early reflections, it was hypothesized that precise details of early reflections could not contribute significantly to room size perception, and early reflections was, hence, left out of this study.

The reverberation characteristics of the room, i.e., the total level of reverberation (or the direct/reverberant sound ratio, henceforth referred to as D/R ratio) and length of the reverberation tail in Fig 1d), are possible cues that have a strong effect on the perception of the size of the room. This is because, in principle, the sound energy stored in a room is a function of the power of the sound source and the volume of the room [4]. The rate at which this energy is absorbed depends on the area of all surfaces and objects in the room as well as their absorption coefficients. In a bare room, the reverberation time is proportional to the ratio of volume to surface [6]

$$T_{60} = 0.161 \frac{\text{Volume}}{\text{Area}}, \quad (1)$$

where  $T_{60}$  is the time it takes for the sound to reach the decay level of  $-60$  dB, and Area stands for the absorption area, which is obtained by multiplying the area of the absorbing surfaces with their absorption coefficients. This means that, in general, large rooms also have longer reverberation times than smaller rooms.

## 2. LISTENING IN NATURAL ENVIRONMENTS

To simulate a speaker speaking in a room, it is necessary to first analyze how a psychoacoustic event occurs in a natural environment. The straight line path between the source and listener, the direct sound, provides the best information about the direction of the source. The direct sound is the least compromised component of the sound and plays an important role in judging the distance as well as direction of the sound source. Sound waves emitted by the source also reflect from objects (in the case of a room, the walls, floor, etc.) and reach the listener, thus constituting the early reflections. Early reflections, therefore, come from fixed directions surrounding the listener in a closed room [4].

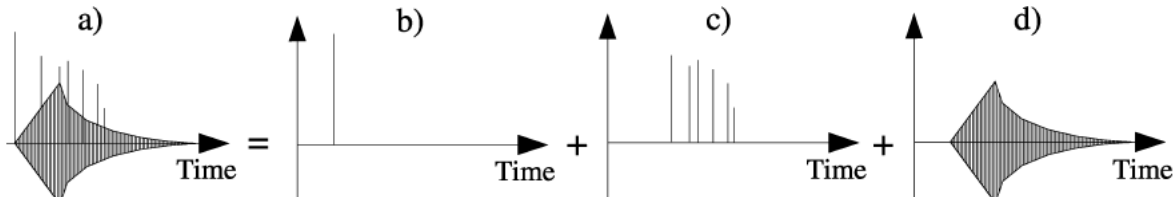


Fig. 1: A room impulse response (a) in the temporal domain consists of three parts. The first peak in the signal presents the *direct sound* (b), that is, the sound which arrives straight from the source to the measuring point. The following few peaks (c) denote the *early reflections*, where the sound waves have been reflected from the surfaces of the room before arriving to the measuring point. As the number of wave reflections, diffraction and diffusion effects increase, it is no longer convenient to simulate each wave separately, and the myriad of reflections can be seen as linearly rising and exponentially decaying noise signal. This part of the impulse response is called the *reverberant part* (d).

In a small room, the early reflections would arrive very close to the direct sound. In larger rooms, they would arrive later in time because of the longer propagation time for the sound waves across the room. Although humans can perceive and are sensitive to the early reflections, they do not affect our ability to localize a sound source in a reverberant environment. This is because the direction of the direct sound (or first wave front) is judged as the direction of the sound source by the auditory system. This effect is called the precedence effect or the law of the first wavefront [2]. As the number of reflections increase, they become closer and closer in time until they merge to form the *reverberant sound*. The reverberation tends to come from all directions and envelop the listener, thereby giving the listener a sense of space. Reverberant energy is a very important acoustic cue and is used by listeners in many mundane tasks in spatial hearing [3]. It affects temporal structure, spectral content, intensity and interaural differences among other parameters. Many characteristics of the source, including relative position of the source in a room, as well as perception of the size of a room can be affected by reverberation.

Previous studies [3] have attempted to investigate the ability of listeners to extract information about the position of the speaker in the room and the position of the listener in the room from the reverberation pattern of the sound. They concluded that listeners are practically unable to reliably identify

differences between different source or listening locations in a room. It was concluded that although listeners are sensitive to gross characteristics of the reverberation, they are relatively insensitive to the pattern of early reflections due to different locations. It can be inferred that changing the listener position or simulated speaker orientation in listening tests should have no significant effect on the results.

### 3. LISTENING TESTS

The listening tests were conducted in an anechoic chamber at HUT. A 16-channel loudspeaker setup was used such that loudspeakers were positioned in different directions around the subject. Besides the horizontal plane around the listeners ears, there were also elevated and lowered speakers.

The first experiment was performed by ten test subjects. They were all male, aged between 22 and 35, and having a general interest in sound and acoustics. The test subjects were briefly interviewed prior to the actual listening test in order to find out possible hearing deficiencies or other things which might bias the test results. The test subjects were only given a brief description of the task they had to perform and no prior information about the test details were given to avoid any effect on the results.

Speech signals were used as source signals, because humans are analytic when listening to them and also because a speech signal probably will not bias the

perceived size of the room in any way, since people are used to hearing speech in various spaces with varying acoustics.

The test followed a paired comparison paradigm. The task was to answer which of the two presented sound samples sounded like a larger room, given that the level of the sound source and the distance between the subject and the source were the same. The subject could hear the samples only once, and the next pair was played after the subject had given his answer. The samples were played in random order so that the correlation between adjacent sample pairs was minimum, that is, the parameter values were not changed gradually.

The subject gave his answer using a cordless computer keyboard. The only source of light in the anechoic chamber was a small pencil light attached to the ceiling and pointed to the keyboard so that using the keyboard was possible without impairing the subject's ability to imagine being in a normal acoustic space instead of the anechoic chamber.

#### 4. SIMULATED ROOMS

The simulated environment used in this experiment was that of a speaker speaking in a room in front of a seated listener. Humans are sensitive to sound from all directions and are able to localize sounds reasonably well in all directions. In rooms, sound reaches the listener not only by the direct path, but also along other paths such as reflections from the walls, floor, etc. In order to produce realistic sound samples, sound must appear from every direction and not just one direction as this would not be natural. This was achieved using a 16-channel speaker setup. The listener was surrounded by loudspeakers, which generate simulations of the direct sound, early reflections and reverberation. Although, because of the limited number of speakers, this scenario is not an exact replica of a natural room, the resolution is good enough as the human directional hearing is not accurate with multiple sounds from different directions [5]. By using uncorrelated noise in the reverberation part of the simulated impulse responses, the loudspeakers also generated a realistic impression of reverberation to the listener sitting at the listening position. Thus a virtual room environment with a speaker talking in front of a seated listener was simulated.

Pseudo-realistic impulse responses were created with the specified parameter values for each loudspeaker. The early reflections and reverberation part were lowpass filtered to simulate attenuation due to propagation in air, as well as reflections from the walls. Frequency shaping was carried out using a cascade of lowpass filters so that the subsequent reflections and reverberations were filtered more than earlier ones.

#### 5. TEST PARAMETERS

The tunable parameter values of the simulated impulse responses were the reverberation time and the amplitude of the reverberation peak. Three different values were used for each parameter, thus yielding a total of  $3^2 = 9$  different sound samples. Four repetitions were carried out for each pair. The total number of trials was 144. The parameter values were chosen by the authors to be as listed in Table 1. The total length of the listening test was approximately 15-20 minutes without breaks.

<i>Parameter</i>	<i>Val 1</i>	<i>Val 2</i>	<i>Val 3</i>
D/R ratio [dB]	-28	-25	-23
Reverberation decay time $T_{60}$ [sec]	0.62	0.73	0.83

Table 1: The parameter values used in the experiment. D/R ratio denotes the amplitude of the reverberation peak w.r.t. the direct sound expressed as decibels. The reverberation decay time is the time for the sound to drop to 60% of its peak value.

#### 6. ANALYSIS

The data was analyzed for significant parameters using two-factor ANOVA. The number of times a sample was chosen as bigger was used as an indicator of how likely that combination of reverberation time and D/R ratio was to be chosen as a larger room over the other eight combinations. The test results, shown in Table 2 revealed that the only significant parameter was reverberation time.

Results showing the effect of reverberation time on listeners' perception of room size are shown in Figure 2.

It can be seen that D/R ratio appears to have almost no effect in the cumulative analysis. Rooms with a

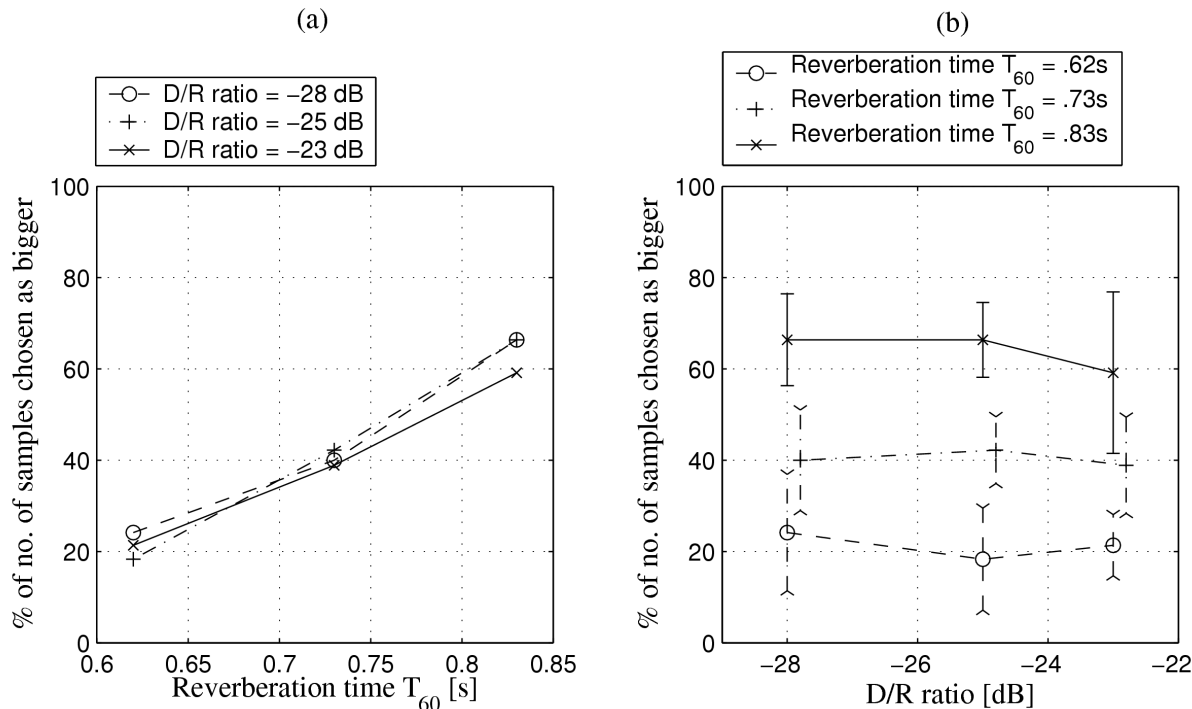


Fig. 2: The number of times a sample was chosen as being bigger (expressed as percentage of the total number of trials in which that sample was presented), at three values of reverberation time, shown for three settings of D/R ratio. The same data is shown in two plots with different x-axis: reverberation time  $T_{60}$  in (a), and D/R ratio in (b). Samples with a larger reverberation time clearly show a greater tendency to be judged as bigger, at all steps of D/R ratio. The points on each plot represent the mean values. The 95% confidence limits for the mean are shown for each point.

<i>Cue</i>	<i>ANOVA p-value</i>
Reverberation time $T_{60}$	$2.82 * 10^{-16}$
D/R ratio	0.639724

Table 2: ANOVA p-values for the two parameters tested in the first experiment. Reverberation time is clearly a significant parameter. D/R ratio is not a significant parameter.

high reverberation time consistently show a greater tendency to be judged as bigger irrespective of the value of D/R ratio. This indicates that reverberation time is the strongest cue in room size perception and easily overrides D/R ratio.

It was observed, however, that there were considerable differences in interpreting D/R ratio among the

test subjects. While some test subjects interpreted a high reverberation level (i.e., lower D/R ratio) as a smaller room, others judged such samples as larger rooms. The performance of each of the ten participants is shown in in Figure 3. It can be seen the the perception of the size of the room is quite different for each listener when only D/R ratio is varied. This implies that D/R ratio is not used in room size estimation.

Physically speaking, smaller rooms tend to have a higher sound pressure level of reverberation as compared to larger rooms. It is easily imagined that sound in a small room, say a bathroom for instance, sounds louder and more reverberant because of the high sound pressure level of reverberation in it. It was hypothesized at the start that D/R ratio should

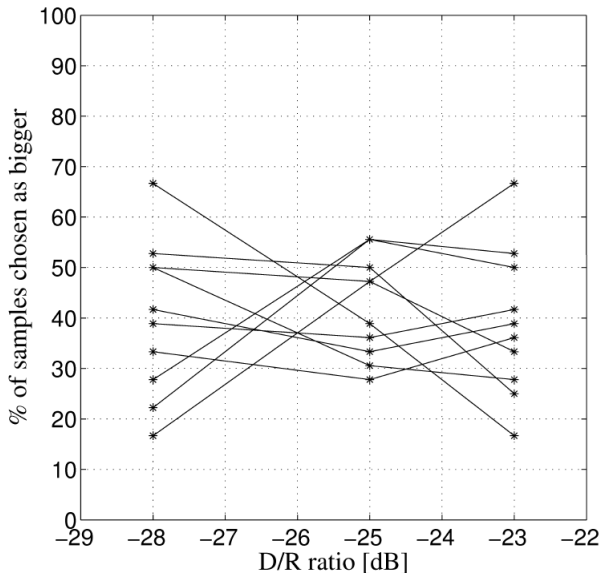


Fig. 3: The variation of number of times a sample was chosen as being bigger (expressed as percentage of the total number of trials in which that sample was presented) by the ten test subjects at three levels of D/R ratio. The results vary quite much, indicating that D/R ratio is not a salient cue in room size perception.

thus be a good cue for room size and was expected to have a strong influence on the listeners judgment. The fact that the listeners are divided in their interpretation of D/R ratio shows that, although it appears to have some influence on the listeners, it is not generally a salient cue for room size perception.

## 7. DISCUSSION AND CONCLUSIONS

It is worthwhile to note that there is no direct relation between the physical size of the room and the reverberation time ( $T_{60}$ ). Sabine's formula predicts the reverberation time in terms of the volume of the room and the area of an absorption window, and is only a general indicator that a large room has a larger reverberation time than a small room. Many factors like the shape of the room, material of the walls and furniture, etc. also contribute significantly to the exact value of  $T_{60}$ . It may be possible for a large room and a medium-sized room to have the same reverberation time. Yet, it appears that, in the absence of other information about the room, the auditory system definitely interprets the information in the acoustics cues of sound in a direct,

reverberation-dependent manner, and greater reverberation time is consistently associated with larger rooms. While this may not strictly be true always, it is true in most cases in the real world. In general, the human auditory system is able to reliably extract information about the room size from the reverberation time alone.

The confusion associated with interpreting D/R ratio by listeners is consistent with earlier studies [1], indicating that D/R ratio is not used as a cue. It is interesting to note that it was assumed that the distance between the speaker and listener was fixed, as a result of which, the direct sound was at the same level in all the sound samples. The sound was also adjusted to be at a natural level, i.e., the level of a normal speaker in a normal room. As a result, a decrease in the D/R ratio would appear to the listener as purely an increase in loudness. Simple loudness or intensity of the sound, therefore, can be taken as a measure of D/R ratio. Some listeners may have (mistakenly) used loudness alone as a subjective measure of room size, specially when the more dominant cue of reverberation time is low.

This would explain the inconsistent effect of D/R ratio on listeners. This only serves to reinforce the conclusion that D/R ratio is not a salient cue for room size estimation, and should not be used as a control parameter in algorithms simulating rooms of controllable sizes.

The aim of the experiment was to assess the relative importance of parameters in judging the size of a room from the point of view of virtual reality and immersive audio applications. The conclusion of the experiment must therefore be that reverberation time is the most dominant cue in room size perception. Reverberation level (D/R ratio) is not used in this task. It is a cue which is interpreted differently by different listeners, and from the point of view of future developments of multimedia or virtual reality algorithms, this parameter should be used with extreme care, if at all, for generating an impression of room size.

## 8. ACKNOWLEDGMENT

This work was conducted in a course on Spatial Sound at HUT. Ville Pulkki has received funding from the Academy of Finland (101339).

## 9. REFERENCES

- [1] J. Sandvad, "Auditory perception of reverberant surroundings", *J. Acoust. Soc. Am.*, vol. 105(2), p.1193, February 1999.
- [2] Blauert J., *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, Cambridge, MA, 1997.
- [3] B. Shinn-Cunningham, "Acoustics and perception of sound in everyday environments", *Proc. 3rd Int. Workshop Spa. Media*, Aisu-Wakamatsu, Japan, March 2003.
- [4] T. D. Rossing, *The Science of Sound* 2nd Ed., Addison-Wesley, 1990.
- [5] Moore B.C.J. ed., *Hearing*, Academic Press, 1995.
- [6] W. C. Sabine, *Collected Papers on Acoustics*, Harvard University Press, Cambridge, MA, 1922.