

# *Linearizing auditory distance estimates by means of virtual acoustics*

Delphine Devallez, Federico Fontana and Davide Rocchesso

## **Abstract**

Auditory distance estimates are not linearly related to physical distances: people tend to overestimate close physical distances, and underestimate far distances [1]. We have modeled a virtual listening environment whose objective is to provide a linear relationship between perceived and physical distance. The simulated environment consists of a trapezoidal membrane with specific absorbing properties at the boundaries. A physics-based model simulates acoustic wave propagation in this virtual space and provides auditory distance cues, namely intensity and direct-to-reverberant energy ratio. Simulations predict the linearity of the psychophysical function relating physical distance to perceived distance, computed from Bronkhorst and Houtgast's model. A listening experiment, involving eleven subjects and using a procedure inspired by the MUSHRA test, shows that the individual psychophysical functions are well described by a linear fit. This promising result suggests that the present virtual environment is a potential tool for providing a direct mapping between perceived and physical distance in auditory interfaces. In addition, the procedure was validated by retesting the same stimuli with the magnitude estimation method as well as reproducing the modified version of the MUSHRA test on a different simulated environment geometry.

## **1 Introduction**

### **1.1 Distance perception in reverberant environments**

Although scientific research has been paying increasing attention to distance perception for the last ten years, auditory displays do not yet take great advantage of this knowledge to provide richer spatial information. Among the acoustic cues used by humans to perceive the distance of a sound source (refer to [2] or [3] for a detailed review), the direct-to-reverberant energy ratio plays an important role in reflective environments. In the far field, the level of reverberation is roughly independent of the source location whereas the intensity level of the direct sound varies inversely with the square of the source distance. As a result the direct-to-reverberant energy ratio decreases with increasing distance. Mershon and King showed that reverberation improves distance perception [4],

and Shinn-Cunningham demonstrated that people judge distance of real sound sources as being greater in reverberant environments compared to anechoic environments [5]. In free-field conditions Gardner [6] and Coleman [7] proved the inability to accurately judge distances, because the most important cue (loudness) is confounded with the level of the sound itself. Nielsen [3] found that reflections are a powerful cue for perceiving distance. Results from their experiments showed that in anechoic conditions there is no correspondence between the physical distance of the sound source and the perceived distance, and judgment was rather based on intensity or timbre changes [7], or influenced by visual cues [6]. Furthermore, reverberation is also a major conveyor of auditory spaciousness and externalization [8]. This is in particular experienced during headphone reproduction, when no natural reverberation is given by the auditory environment.

Of major interest are research studies conducted on distance perception in reverberant spaces which report that distance estimates are not linearly related to physical distances: people usually underestimate far away source distances and overestimate source distances that are closer than 1 m [2]. This is an important result because it shows that the perception of auditory distance is distorted compared to the physical distance. By carrying out listening experiments and using data from numerous previous studies, Zahorik suggested the use of a compressive power function of the form  $r' = kr^a$ , where  $r'$  is the estimate of the perceived distance and  $r$  is the physical source distance, to model the relationship between perceived and physical distances. Furthermore the author showed that this model was a good fit to 84 data sets on distance perception [1]. From a practical point of view, it means that the perceived distance may be deduced from that psychophysical function, once the physical distance is known.

## 1.2 Bronkhorst and Houtgast's model

Another model, proposed by Bronkhorst and Houtgast [9], has shown to converge with other research results used by Zahorik in its power function fit analysis [1]. Bronkhorst and Houtgast's model relates the estimate of perceived distance to the reverberation cue. The amount of reverberation is commonly assessed by measuring the ratio between the direct sound energy and the reverberant sound energy. More precisely, the direct sound energy is generally considered to lie within the first 2.5 ms of the sound signal. This time window was chosen because it approximates the duration of Head Related Impulse Responses measured in anechoic conditions and therefore captures the direct path of the sound signal [10]. Bronkhorst and Houtgast [9] proposed the use of a larger integration window (6 ms) for determining the energy of the direct signal. This choice is supported by the observation that listeners underestimate far physical distances (effect called "auditory horizon"). By increasing the window size, the decrease of the direct sound energy with distance has less effect, meaning that the overall energy contained in the window decreases more slowly. Furthermore, the choice of 6 ms is in line with studies conducted on the *precedence effect* involved in localization tasks. Indeed Houtgast and Aoki [11]

showed the dominance of the first 5 to 10 ms of the sound signal in cues used in sound localization. These values may also be related to the 8 ms temporal resolution of the auditory system, found by Moore et al. [12]. Coming back to Bronkhorst and Houtgast's model, the integration time window  $W$  is defined as:

$$W = \begin{cases} 1 & \text{for } 0 < t \leq t_W - \frac{\pi}{2s} \\ 0.5 - 0.5 \sin[s(t - t_W)] & \text{for } t_W - \frac{\pi}{2s} < t < t_W + \frac{\pi}{2s} \\ 0 & \text{for } t \geq t_W + \frac{\pi}{2s} \end{cases}$$

with  $t_W \simeq 6.1$  ms and  $s \simeq 400$  s<sup>-1</sup>. Finally, the perceived distance  $d_s$  of a sound source is expressed as:

$$d_s = Ar_h \left( \frac{\hat{E}_r}{\hat{E}_d} \right)^j \quad (1)$$

where  $r_h$  represents the critical distance (also called reverberation radius),  $A$  and  $j$  are parameters determined empirically through experiments,  $\hat{E}_r$  is the modified energy of the direct sound calculated using the integration window  $W$ , and  $\hat{E}_d$  is the modified energy of the reverberant sound calculated using the integration window  $1 - W$ .

Thanks to this quantitative model, it is possible to predict the perceived distance of a sound source in a given acoustical space. We will make use of the same model for synthesis rather than analysis purposes, i.e. to shape a virtual acoustical environment in order to render sound source distances.

### 1.3 Objective

As seen in section 1.1, the relationship between perceived distance and physical distance in a real reverberant room is not linear but rather logarithmic. Our question is: *Can we create a virtual environment in which the psychophysical function between perceived and physical distances is linear?* Creating virtual auditory environments may indeed offer new opportunities to manipulate auditory distance cues. Besides, the model proposed by Bronkhorst and Houtgast may assist the design of the environment by predicting the perceived distance of a sound source based on the computation of the corresponding room impulse response. From a practical point of view, such a virtual environment may become a tool for interactive sound spatialization. The ability to manipulate the relative depth of sound sources and create dynamic distance effects may enrich the content and quality of auditory displays, or offer novel human-computer interactions.

In the field of sound spatialization, recent research has targeted the physical modeling of acoustic propagation. In particular, the use of a physics-based approach such as digital waveguide modeling, is motivated by the simple formulation of the model and its intuitive control with parameters that have a physical meaning. Fontana and Rocchesso [13] have underlined the effectiveness of a Digital Waveguide Mesh (DWM) modeling a rectangular parallelepiped to provide acoustic depth cues. Listening experiments using this model have shown

its ability to render the apparent distance of sound sources, and results were consistent with previous studies of Zahorik [2] and Bronkhorst and Houtgast [9]. The peculiarity of this virtual environment is the exaggeration of reverberation. Since listeners tend to underestimate far physical distances, increasing the amount of reverberation may help reduce the compression of the perceived distance with respect to the physical one. Besides the acoustic features of the proposed environment, the objective of this research is not to recreate a true-like listening scenario such as those obtained, for instance, by using auralization techniques based on Head-Related Transfer Functions. Rather, we want to design a spatialization tool that provides a representation of the same scenario that is immediately apparent to most listeners, and can be reproduced by most consumer systems even under non-ideal listening conditions.

In the next section we present a spatialization model using the DWM which is devoted to render the sound source distance from the listener, and in particular to provide a linear mapping between the physical distance of the source in the virtual environment and the estimate of the distance perceived by the listener.

## 2 The trapezoidal membrane

After some informal experimentation and assessment of the reverberation cue provided by several 3D shapes, especially pyramids, our attention was finally captured by a 2D vibrating surface (i.e. a membrane), whose shape resembles that of a trapezoid. Such simulations have been conducted by varying the shape and boundary absorptions in a DWM modeling the environment under analysis. In the following we provide a summary of the main features of our DWM model.

### 2.1 Geometry of the DWM

In recent years, it has become possible to model complex, large-scale acoustical environments by means of the DWM [14]. Besides, this method has been shown to be accurate especially at low frequencies [15], and may be used to simulate real-world acoustical rooms with an adequate modeling of the boundaries by means of properly tuned absorption filters ([16], [17]). The DWM and its different geometries have largely been described in past literature, see [18], [15] and [19]. Our sound propagation model consists of a two-dimensional rectilinear DWM having a trapezoid shape as shown in Figure 1. Each internal junction is connected to four other junctions via waveguides providing acoustic wave transmission. Each waveguide models the wave decomposition of a pressure signal  $p$  into its wave components  $p^+$  and  $p^-$ , and each lossless junction scatters 4 input signals coming from orthogonal directions,  $p_1^+$ , ...,  $p_4^+$  into corresponding output signals  $p_1^-$ , ...,  $p_4^-$  (see Figure 2). The digital waveguide theory states that the sampling frequency  $f_s$  of a bidimensional mesh is determined as [19]:

$$f_s = \frac{c\sqrt{2}}{d_W} \quad (2)$$

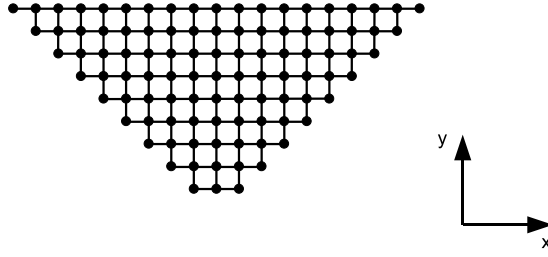


Figure 1: *The two-dimensional rectilinear waveguide mesh representing a trapezoid geometry.*

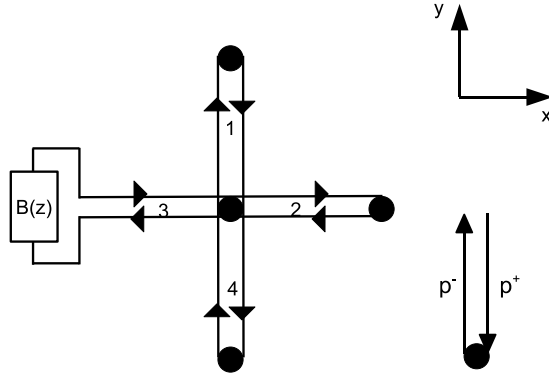


Figure 2: *Zoom on a junction which is connected to other junctions via waveguides 1, 2 and 4. Waveguide 3 leads to a partially absorbing section, modeled using a digital waveguide filter. Triangles filled in black represent oriented unit delays.*

where  $c$  is the nominal wave propagation speed and  $d_W$  the spatial sampling interval or the length of each waveguide. Note that this relation is exact only for diagonal propagation and at  $f = 0$  because of the dispersion error (some spatial frequencies travel slower along the mesh), however this effect may be neglected for the present application. From equation (2) we can deduce the length  $d_W$  of each waveguide. For  $c = 343 \text{ m/s}$  the speed of sound in air and  $f_s = 44.1 \text{ kHz}$ , a waveguide corresponds to the propagation of sound pressure waves on a distance of 1.10 cm. In order to reproduce distances up to 10 meters, the total length of the mesh along the y-axis (see Figure 1) is chosen to be 10.05 m, which amounts to 914 waveguides. Since the mesh is symmetric, the top base of the trapezoid equals  $10.05 \times 2 = 20.1 \text{ m}$ .

## 2.2 Boundary conditions

The properties of the wall materials contribute to the acoustics of a space. This is also the case for a bidimensional acoustic environment since horizontal waves interact with the surface's boundaries. Reflections from surfaces are modeled by Digital Waveguide Filters (DWF), whose coefficients have been tuned to model specific reflective properties of surfaces [18]. The conversion from measurements of physical parameters of real surfaces to DWF coefficients follows the method described in [16]. For a given material, its absorption coefficients measured for octave bands (125, 250, 500, 1000, 2000 and 4000 Hz) are converted to the reflection transfer function using the following relationship:

$$|R(j\omega)| = \sqrt{1 - \alpha(\omega)} \quad (3)$$

where  $R(j\omega)$  is the reflection transfer function and  $\alpha(\omega)$  are the frequency dependent absorption coefficients. Then the `invfreqz` function in `Matlab` has been used to design a stable first-order IIR filter fitting the data.

The boundaries of our trapezoid DWM exhibit the following properties (refer to Figure 3 for the definition of the surfaces  $A$ ,  $B$  and  $C$ ):

- Surfaces  $A$  and  $B$  provide the simulation of hard surfaces, such as bricks or plaster, using the following absorption coefficients for the octave bands defined earlier: 0.02, 0.02, 0.03, 0.03, 0.04 and 0.05.
- Each surface of the group  $C$  is modeled to behave as total absorber. We chose to cancel reflections from these panels in order to avoid audible echoes created from waves traveling back and forth between surfaces  $A$  and  $C$ .

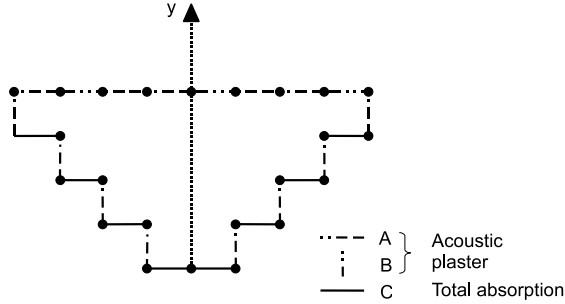


Figure 3: *Properties of the mesh boundaries.*

### 3 Acoustics of the trapezoidal membrane

#### 3.1 Listening and sound source locations

The listener and the sound source are assumed to be point-wise and are positioned on one of the mesh junctions on the y-axis. The listening point is located at the junction no. 2 near the smaller base of the trapezoid, while the sound source is successively positioned at junctions no. 93, 184, 275, 366, 457, 547, 638, 729, 820 and 911 in order to simulate one-meter steps ranging from 1 to 10 meters away from the listening point.

#### 3.2 Acoustical properties of the listening environment

Simulations of the listening environment for the ten sound source positions were carried out in `Matlab`, and ten corresponding impulse responses were computed. Figure 4 shows the impulse responses measured at 1 and 6 m, and Figure 5 shows the frequency responses at the same distances, up to 5 kHz.

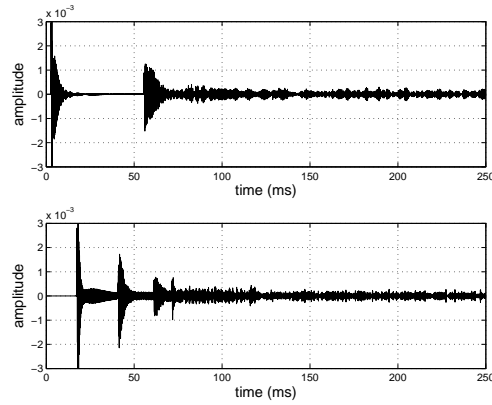


Figure 4: *Impulse responses inside the trapezoid mesh. Top: 1 m. Bottom: 6 m.*

##### 3.2.1 Overall intensity level

Among the prominent cues for auditory distance perception, intensity is the most noticeable: when the sound source moves away from the listener, intensity decreases. However, the level of direct sound varies both with distance and with the energy emitted from the sound source, so that the listener needs some a priori knowledge about the sound source level in order to evaluate the egocentric distance. Another drawback of using the intensity cue for judging distance is the unreliability of this piece of information for auditory displays, since it can be manipulated by users in most interfaces. Figure 6 shows how the total energy

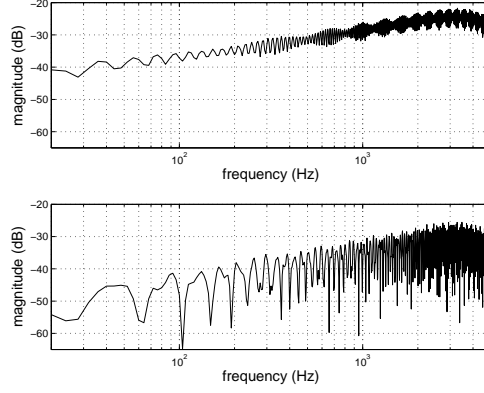


Figure 5: *Frequency responses inside the trapezoid mesh. Top: 1 m. Bottom: 6 m.*

of the impulse responses varies with distance. The energy inside the trapezoid mesh decreases significantly less than in open space, characterized by the well-known 6 dB law. The present 2D trapezoid DWM exhibits a decrease of only

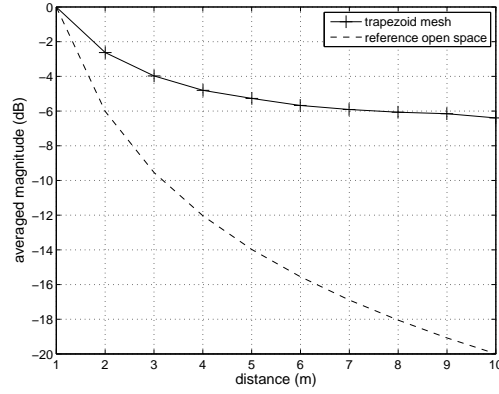


Figure 6: *Average magnitude of the impulse response as a function of distance.*

6 dB from one to ten meters, in particular the intensity change over the range 6–10 m is less than known just-noticeable differences for intensity ( $\sim 1$  dB) [20]. In that distance range, it may be concluded that the overall intensity of the sound signal may hardly provide a reliable source of information to judge distance.



### 3.2.2 Direct-to-reverberant energy ratio

The virtual acoustic space we have designed aims at rendering depth information mainly through the direct-to-reverberant energy ratio cue. Figure 7 shows the values of the direct-to-reverberant energy ratios for the ten distances in the trapezoidal mesh. The method used to compute these ratios is the following:

1. For each impulse response the delay of the direct sound is deduced from the distance between the sound source and the listening point. It is then removed from the impulse response.
2. The direct energy is integrated among the usual first 2.5 ms of the delay-free impulse responses.
3. The reverberant energy is calculated from the tail of the delay-free impulse responses.

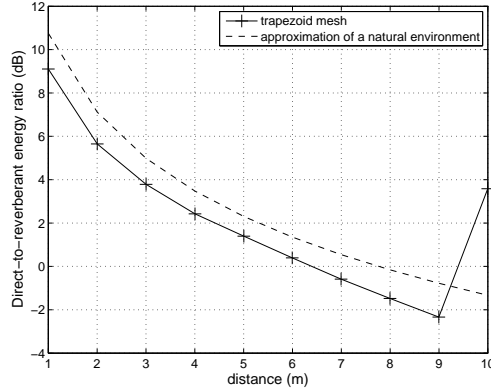


Figure 7: *Direct-to-reverberant energy ratio as a function of distance.*

For comparison the direct-to-reverberant energy ratio  $v$  was computed for a natural environment, modeled by Zahorik with the function  $v = -3.64 \log_2(r) + 10.76$  [21]. The two curves follow the same trajectory for distances up to 9 m, suggesting that on this distance range the direct-to-reverberant energy ratio in the virtual environment follows a natural behavior. Numerical artifacts in Digital Waveguide Mesh simulations, which affect the responses especially in the proximity of the mesh boundaries, may be responsible for the increase of the direct-to-reverberant energy ratio between 9 and 10 m.

The other important auditory distance cues, namely the binaural cues and the spectrum, are not provided by the present model. First, binaural cues, which arise due to differences between the two ears, are only relevant if one wishes to reproduce sound sources outside the median plane, and require two

different reproduction channels. Besides, past studies do not agree on their utility for perceiving distance, in particular in the far-field [1]. About the spectrum changes due to the air attenuation of high frequencies, they occur for very large distances (superior to 15 m) which lie beyond the available range of our simulated environment. To summarize, the only pieces of information about the distance of a sound source provided by the proposed model are the intensity and mostly the direct-to-reverberant energy ratio.

### 3.3 Application of Bronkhorst and Houtgast’s model

To apply Bronkhorst and Houtgast’s model, the critical distance  $r_h$  is to be estimated. It corresponds to the distance at which the sound pressure level of the direct and the reverberant sound fields are equal, providing a direct-to-reverberant energy ratio of 0 dB. From Figure 7, this condition arises at a distance of about 6 m. A more accurate computation of the direct and reverberant energies focused around this position leads to a value of 6.3 m. Finally, the modified direct-to-reverberant energy ratios for the ten distances are calculated from eq. 1 and are shown in Figure 8. Note that the explanation given for the increase of the direct-to-reverberant energy ratio at the far end of the membrane also applies to the behavior of the distance estimate at the greatest distances. Before the analysis, it should be noted that what is

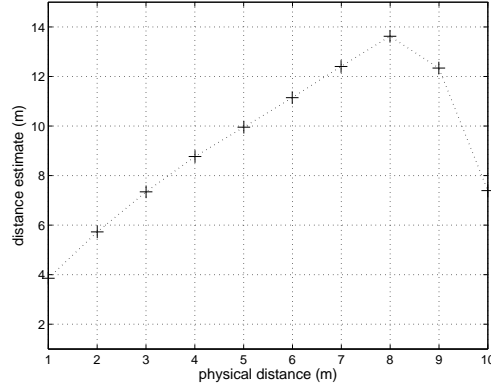


Figure 8: *Distance estimate according to Bronkhorst and Houtgast’s model.*

called *physical distance* on this plot relates in fact to a modeled distance inside the mesh derived from the digital waveguide theory. An almost linear behavior arises for physical distances from 1 to 8 m along the mesh. This behavior means that Bronkhorst and Houtgast’s model predicts that in the trapezoidal mesh the perceived distance is a linear function of the physical distance in that range. In other words, listeners should perceive as identical the distance changes between every couple of adjacent physical distances in the range 1-8 m. Moreover, the

uncertainty of some parameters values used in the calculation of Bronkhorst and Houtgast’s model, such as the directivity factor and other empirical coefficients, as well as the theoretical definition of physical distance inside the artificial mesh, do not allow to determine the slope of the psychometric function. This means that the range of perceived distances cannot be established.

## 4 Experiment I

### 4.1 Subjective assessment of the distance linearity

Common procedures used to estimate psychophysical distance functions include magnitude estimation (on explicit scales such as meters or on dimensionless scales), or other methods such as perceptually directed action [22] consisting of walking to the perceived location. They generally aim at evaluating the absolute egocentric distance of a sound source, one at a time, and may suffer from adaptation effects. Subjects may indeed readjust their responses due to the subsequent stimuli as they settle on a range of scale values. The objective of our listening test consists of assessing the linearity of distance perception, therefore we are interested in evaluating relative distance perception of a set of sound sources. Considering possible bias due to learning effect and high variability of responses across trials, we thought that the common evaluation methods were not adequate for the present issue. Preferably a method allowing direct comparison of the different stimuli would provide a more robust estimate of the relative distance perception of a set of sound sources. Such a methodology, the MUSHRA test, exists for the subjective evaluation of audio quality, and is defined by ITU-R recommendation BS.1534-1 [23]. MUSHRA stands for MULTiple Stimuli with Hidden Reference and Anchor. This methodology was developed for the subjective evaluation of audio coding systems and allows the comparison of high quality reference sounds with several lower quality test sounds. The aim of each experiment is to compare a high quality reference sound to several test sounds sorted in random order, including the reference. According to the procedure, each subject is asked to assess the quality of each test sound (relative to the reference and other test sounds) by grading it on a quality scale between 0 and 100. It is not necessary that one test sound be graded 0, however at least one must be graded 100 (because the reference sound is among the test sounds). The test sounds must also include one or several *anchor sounds* computed similarly for all experiments using simple signal processing operations. In particular, one of these anchor sounds must be the reference sound low-pass filtered at 3.5 kHz.

### 4.2 Method

#### 4.2.1 Participants

Eleven Italian volunteers (4 women and 7 men), studying or working at the University of Verona, Italy (except one teenager aged 15), aged 15–41 years old,

participated in the experiment. All reported to have normal hearing.

#### 4.2.2 Stimuli and Apparatus

The objective of the listening test was to assess the linearity of distance perception for physical distances ranging from 1 to 8 m in the trapezoid mesh, as predicted by Bronkhorst and Houtgast’s model. The 8 corresponding impulse responses, computed at 44.1 kHz and 0.25 s long, were convolved with a 4 s speech signal. The latter consisted of a female voice saying the words “Kroklok-wafzi? Sememememi” (opening of Christian Morgenstern’s poem *Das grosse Lalula*) already used in a previous study [24]. The stimuli were generated by an interactive `Matlab` script and presented over an AKG K240 open headphone set. Subjects gave their answers on a computer terminal with a numeric keypad.

#### 4.2.3 Procedure

Distance judgments were collected using a modified version of the MUSHRA test [23] with hidden reference. This reference was the sound source located at 1 m in the virtual environment. Prior to the experiment, listeners received a written instruction explaining the two phases of the test. First a training phase allowed the user to browse through the 8 stimuli, ordered randomly on the computer screen, and get acquainted with the auditory distance range. In the second phase, namely the evaluation phase, subjects had to rate the perceived distance of the 8 stimuli against the reference sound, on a scale ranging from 0 to 10 dimensionless units, through steps of 0.1 units. Each sound was graded by moving the corresponding slider on the computer screen, and the grading value appeared in the corresponding text field below. Again, the 8 stimuli were randomly ordered for each trial. The hidden reference was also included in the test, and listeners were instructed to rate this particular sound to the value “1”. Besides, they were instructed to use the value “0” for sound sources that appeared to originate from their head. They could listen to the signals in any order, as many times as they wanted. When they were satisfied with their grading of all signals, they clicked on the “save and proceed” button on the screen to go to the next trial. In total each subject had to compute the task five times, which took approximately 15–20 minutes. Figure 9 shows a snapshot of the user interface. The whole procedure was carried out on `Matlab` using a modified version of the MUSHRAM package [25], a set of `Matlab` routines dedicated to MUSHRA listening tests. In both phases of the experiment, the sequence of the stimuli was randomized.

### 4.3 Results

Results for each subject are displayed in Figure 10. None of the participants used the value “0”. However at this stage, we cannot conclude that they perceived all the simulated sound sources as external to the head since they had to grade the closest sound source a “1”.

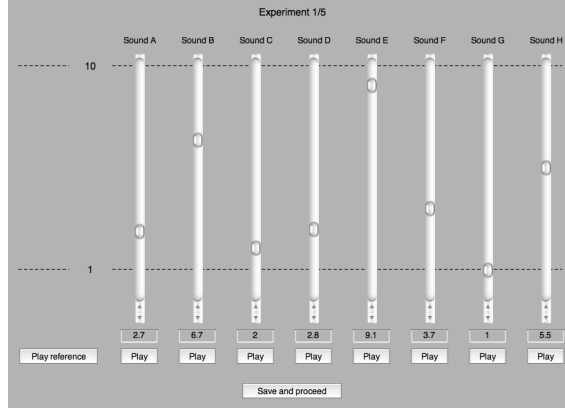


Figure 9: *Snapshot of the user interface during the evaluation phase.*

The observation of the individual psychometric functions on the whole range of sound source distances does not reveal evident similarities. An analysis of the results for each subject shows that the individual relationships between perceived and physical distance are better described with linear functions [ $r^2 = 0.7071$ ,  $F(1, 38) = 91.73$  and  $p = 1.117 \times 10^{-11}$  for the worst fit] than with power functions [ $r^2 = 0.5894$ ,  $F(1, 38) = 54.56$  and  $p = 7.424 \times 10^{-9}$  for the same data]. Power functions are of the form  $r' = kr^a$ , where  $r'$  is an estimate of perceived distance, and  $r$  is the physical distance of the sound source. Parameters of each individual fitted linear and power functions are reported in Table 1 which also shows the goodness-of-fit ( $r^2$ ) for each fit. Figure 10 suggests that linearity is more apparent for near sound source distances. In order to assess more precisely this indication, a linear regression was applied to each individual psychometric function for 5 different distance ranges, namely 1-8 m, 1-7 m, 1-6 m, 1-5 m and 1-4 m. The box plot of Figure 11 summarizes the distribution of the standard errors of the slope for each distance range. By taking the mean of the individual standard errors of the slope of the linear regressions for each distance range, the minimum is found for the whole distance range of 1 to 8 meters (mean standard error  $< 0.072$ ). The analysis of the best linear fit for this distance range gives the following results:  $r^2 = 0.8946$ ,  $F(1, 38) = 322.6$  and  $p < 2.2 \times 10^{-16}$ . From these results we may conclude that the relation between subjective and physical distances in the proposed virtual environment is well described with a linear function.

#### 4.4 Test Repeatability

Since the modified version of the MUSHRA test is not a common procedure in auditory distance perception, we first propose to check the repeatability of the

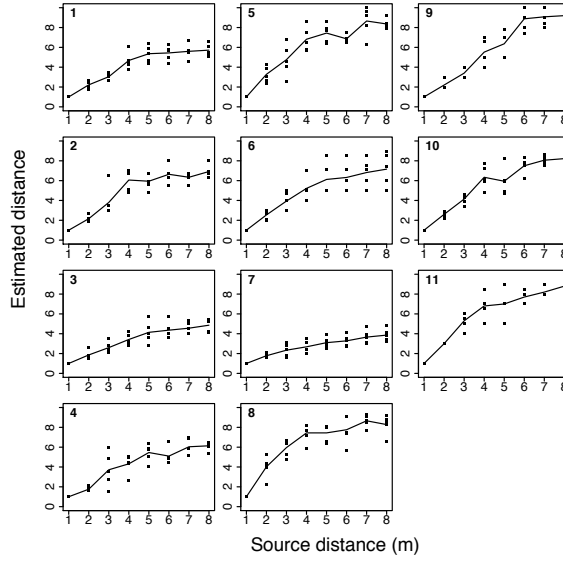


Figure 10: *Exp. I: Individual apparent distance judgments as a function of physical distance. Each plot displays data for one subject. Means are connected by a line.*

test, i.e. the consistency of individual responses over repetitions of the same experiment under the same conditions. This may be easily done by estimating the standard deviation of the five repeated evaluation tasks for each subject. An analysis of variance was performed with the number of repetitions as a within-subjects factor. No statistically significant differences between the different sessions were found, meaning that the procedure is repeatable [Mean square= 2.71,  $F = 1.1724$ ,  $p = 0.2795$ ].

## 5 Experiment II

In order to validate the scaling procedure described in Section 4, we propose to compare it to another procedure commonly used to assess distance perception. Although it was shown that all common procedures produce similar results [1], the magnitude estimation procedure was chosen since it was used in a reference study [26].

### 5.1 Method

The same eleven volunteers as for experiment I participated in the second experiment. The stimuli and apparatus were those already used in experiment I. Only the procedure differed: after listening to the 8 stimuli ordered according

Table 1: Exp. I: Individual fitted linear and power function parameters for each subject.

subject n.	Linear fit			Power fit		
	slope	intercept	$r^2$	$a$	$k$	$r^2$
1	0.69	1.04	0.78	0.86	1.15	0.74
2	0.84	1.07	0.74	0.96	1.15	0.72
3	0.55	0.86	0.78	0.77	1.06	0.77
4	0.75	0.84	0.74	0.92	1.05	0.73
5	1.02	1.33	0.77	0.99	1.33	0.68
6	0.86	1.00	0.72	0.93	1.21	0.71
7	0.39	0.96	0.75	0.63	1.08	0.75
8	0.95	2.06	0.71	0.95	1.51	0.59
9	1.30	-0.14	0.90	1.13	1.02	0.85
10	1.04	0.79	0.86	1.01	1.20	0.77
11	1.05	1.26	0.83	1.01	1.32	0.70

to increasing distance, the listeners were asked to judge the distance of the 8 different stimuli presented in a random order via a unitless magnitude estimation procedure. Subjects were free to choose their own standard, assigning any number to the first stimulus and all subsequent ones. Responses were made after listening to each stimulus on the computer terminal with the numeric keypad, and participants could use a precision of one decimal. As in experiment I, a grade of “0” was used for sound sources that appeared to originate from their head. Subjects had to perform the task 3 times, leading to a total of 24 trials.

## 5.2 Results

Figure 12 shows the results for each subject. A grade of “0” was used by five subjects for rating the 1 m stimulus (subject 3 also rated once the 2 m stimulus to “0”). This did not happen in experiment I, because the participants were constrained to use the grade “1” for the closest sound source corresponding to the reference source (therefore being easily identified). If we discard the zero attributed by subject 3 to the 2 m sound source, we may deduce that some listeners may hear the closest sound as originating from their head. Nevertheless we think that this fact does not prevent from applying the model to an auditory interface, as long as users can discriminate between sound sources located at different distances.

Except for subjects 1 and 8, the relation between perceived and physical distance is better described with a linear function [ $r^2 = 0.4778$ ,  $F(1, 22) = 20.13$  and  $p = 1.838 \times 10^{-4}$  for the worst fit] than with a power function [ $r^2 = 0.393$ ,  $F(1, 22) = 14.24$  and  $p = 1.045 \times 10^{-3}$  for the same data]. These data agree with those of experiment I, and parameters of each individual fitted linear

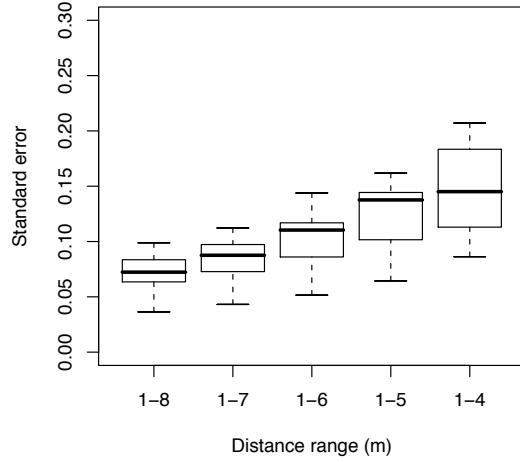


Figure 11: *Exp. I: Distributions of the individual standard errors of linear regressions applied to 5 distance ranges.*

and power functions may be seen in Table 2. For subjects 1 and 8, the logarithmic fit gives slightly better results, however the data for subject 8 are still not well approximated by a power function since the variance accounted for is low [ $r^2 = 0.4276$ ,  $F(1, 22) = 16.43$  and  $p = 5.29 \times 10^{-4}$ ]. It can also be seen from Figure 12 that subject 8 might have reversed the distance scale for some judgments. Besides, his/her results are not consistent through the three repetitions (An analysis of variance with the number of repetitions as a factor gives: Mean square= 18.2756,  $F = 28.724$ ,  $p = 2.578 \times 10^{-5}$ ). Subject 8 is therefore discarded in the following.

By comparing the shape of the psychophysical functions Figure 10 and Figure 12, data of many subjects look qualitatively similar for both response procedures, except for subjects 3 and 7. However, by looking at the individual responses, data in Figure 12 seem to be more scattered. In order to quantify this difference, standard scores were calculated since both experiments used different scales (in experiment II the scale was free whereas in experiment I subjects could rate the perceived distance up to 10). More precisely, the absolute value of the z-scores was calculated for each value of distance estimate, each experiment, and each of the remaining 10 subjects. Then the mean was calculated for each subject and for each experiment. Figure 13 shows the results. It can be seen from this plot that the dispersion is greater in experiment II, and this is validated by calculating the mean of the absolute values of the z-scores over all participants (2.50 for experiment I, and 4.65



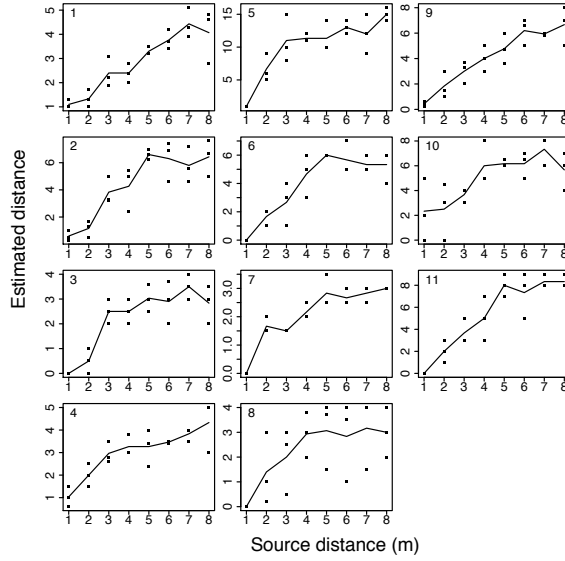


Figure 12: *Exp. II: Individual apparent distance judgments with the magnitude estimation procedure. Each plot displays data for one subject. Means are connected by a line.*

for experiment II). The high variability in distance judgment for a fixed sound source distance is well known [1], and it seems that the procedure derived by the MUSHRA test in experiment I allows to reduce it and provide more consistent responses.

## 6 Experiment III

The linearity observed in the proposed virtual environment with the procedure used in experiment I may be validated by carrying out the experiment on stimuli spatialized in a different virtual environment in which the linearity of the psychophysical functions is not expected. For this purpose, a 2D rectangular DWM is proposed as the comparison environment, and has the following characteristics (See Figure 14):

- Dimensions:  $5.28 \times 0.31 \text{ m}^2$
- Boundary conditions: reflections from surfaces are modeled by first-order digital filters which simulate
  1. reflective surfaces A with a reflection coefficient equal to 0.95
  2. surfaces B as total reflectors

Table 2: Exp. II: Individual fitted linear and power function parameters for each subject.

subject n.	Linear fit			Power fit		
	slope	intercept	$r^2$	$a$	$k$	$r^2$
1	0.49	0.64	0.81	0.49	1.91	0.82
2	0.88	0.43	0.70	0.82	1.56	0.69
3	0.44	0.26	0.62	0.73	1.09	0.61
4	0.40	1.21	0.70	0.43	2.16	0.66
5	1.56	3.15	0.65	0.88	3.01	0.56
6	0.79	0.38	0.66	0.92	1.24	0.65
7	0.37	0.42	0.75	0.61	1.28	0.65
8	0.39	0.56	0.37	0.65	1.23	0.43
9	0.89	0.11	0.83	0.82	1.52	0.78
10	0.66	2.02	0.48	0.59	2.44	0.39
11	1.24	-0.24	0.80	1.09	1.22	0.74

Please note that the length of the new environment is about half the length of the trapezoid, thus we expect that listeners will use smaller values to rate the estimates of the perceived distances. Figures 15 and 16 show respectively the intensity and the direct-to-reverberant energy ratio cues in the rectangular DWM. While the intensity decrease is more linear than in the trapezoid mesh (Compare Figures 6 and 15), the direct-to-reverberant energy ratio shows a sudden decrease in the beginning of the distance range and remains almost constant for distances greater than about 2 m.

Applying Bronkhorst and Houtgast’s model to this model leads to the psychophysical function shown in Figure 15. The resulting curve is not linear and is better described by an exponential function: it seems that this particular virtual environment should produce an expansion of perceived distance, in opposite to the compression observed in most real-world listening environments.

## 6.1 Method

This experiment was run on the same subjects as for experiments I and II. Apparatus and procedure were identical to those used in experiment I. Only the spatialization of the stimuli was different. The 4 s speech signal used for experiments I and II was convolved with 8 impulse responses computed every half-meter from 0.5 to 4 m.

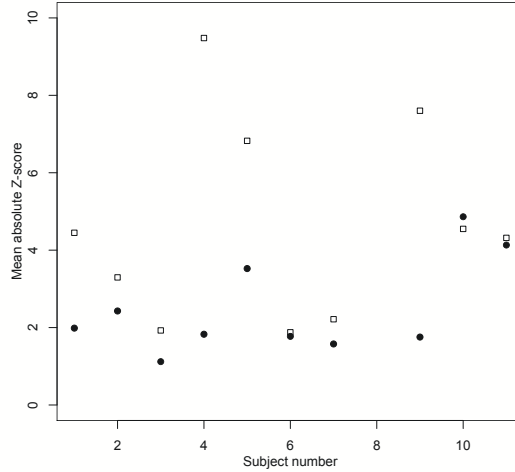


Figure 13: *Exp. II: Mean absolute z-score as a function of subject number. Solid circles: experiment I. Blank squares: experiment II.*

## 6.2 Results

Figure 16 shows the estimates of perceived distance in the rectangular DWM for each subject. A comparison with the scales used by the same participants in experiment I (Figure 11) shows that the distances are effectively estimated as closer. An analysis of the results for each subject was carried out by fitting the estimates with a linear fit, a logarithmic fit and an exponential fit. Except for subjects 2 and 8, the relation between perceived and physical distance is better described with a power function, however the variance accounted for is quite low for some subjects [ $r^2 = 0.352$  for subject 5]. Estimates given by subjects 2 and 8 are rather fitted by a linear regression but again with a quite low variance accounted for [ $r^2 = 0.6542$  for subject 2, and  $r^2 = 0.554$  for subject 8] (See Table 3 for details on the fitted function parameters). As a consequence, it is rather difficult to draw conclusions from the analysis. It seems however that for most subjects the psychophysical functions are not linear, meaning that unlike the trapezoidal membrane, the virtual auditory environment modeled by a rectangular DWM does not produce linear auditory distance estimates.

## 7 Conclusions

We have presented a trapezoidal DWM as a virtual environment able to render distance cues. According to the distance perception model of Bronkhorst and

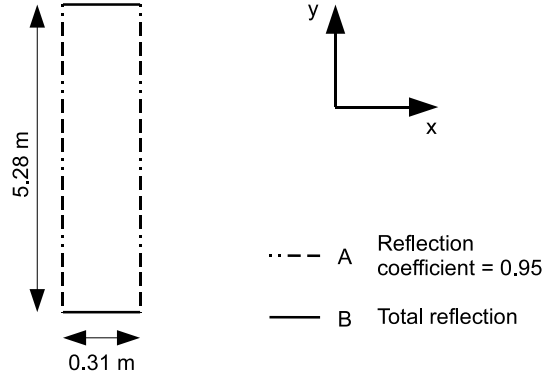


Figure 14: *The two-dimensional rectangular waveguide mesh.*

Houtgast, this environment allows a linear scaling between perceived and physical distance. An experiment was conducted to assess the model, with stimuli in the 1-8 m range. Results reveal that the trapezoidal membrane can effectively render linear distance perception in that distance range.

The procedure used in the experiment derives from the MUSHRA test and was proposed as a method suitable for judging the relative distance of multiple sound sources. In a second experiment the direct scaling procedure was applied to the same stimuli. Results allowed to validate the modified version of the MUSHRA test as well as to reveal the lower variability of responses across trials with the latter method. Finally, a third experiment was conducted using again the modified version of the MUSHRA test on stimuli spatialized in a rectangular DWM and ranging from 0.5 to 4 m. Participants accordingly perceived the sound source distances as being shorter. Besides, Bronkhorst and Houtgast's model did not predict a linear psychophysical function, and this was verified by an analysis of the individual responses.

The linearity of the trapezoidal membrane was assessed for only one distance range (1–8 m), and one may argue that the effect might be dependent on the distance range under consideration. One answer to this issue is the recent study by Zahorik [27] which shows that for most subjects, apparent distance biases do not depend on the range of source distances presented within a block of trials. Since no externalization technique was applied to the stimuli, it may be questionable as to whether the simulated sound sources are perceived as external to the head. It seems that individual Head-Related Impulse Responses lead to the largest externalization [28]. However, to the authors knowledge, past literature has never proposed auditory distance experiments comparing stimuli providing only intensity and reverberation cues versus spatialized stimuli containing additional binaural information, in terms of sound source externalization in the median plane. Moreover, the effect of binaural cues on distance perception of sound sources in the median plane is rather limited, at least in the far-field [26].

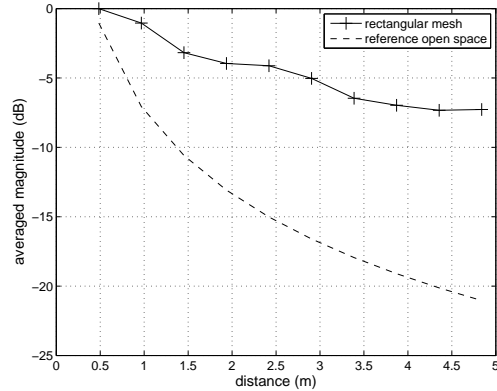


Figure 15: *Average magnitude of the impulse response as a function of distance in the rectangular mesh.*

Furthermore, in the experiments subjects seldom graded the stimuli a “0” which was reserved for those sound sources that appeared to originate from a location within their head. In fact, only the closest sound source received a grade of “0” a few times. In our opinion, externalization is not an issue in the current context as long as listeners are able to discriminate different sound source distances. Although the realism of the simulated acoustic environment would ask for employing more sophisticated acoustic rendering techniques, the proposed spatialization technique allows to render depth with a monaural signal and does not rely on the reproduction hardware configuration, unlike techniques using Head-Related Transfer Functions.

Finally, virtual environments designed for rendering auditory distance may offer novel tools for interactive sound spatialization. In particular, a linear mapping between perceived and physical distance may enable a more transparent user manipulation of sounding objects displayed along the depth dimension and thereby improve the perceived quality of interaction. We believe that auditory perspective effects combined with direct manipulation can become a powerful tool for future auditory interfaces.

## 8 Acknowledgments

This research work has been supported by the European project FP6-NEST-29085 CLOSED - Closing the Loop of Sound Evaluation and Design under the path “Measuring the impossible”.

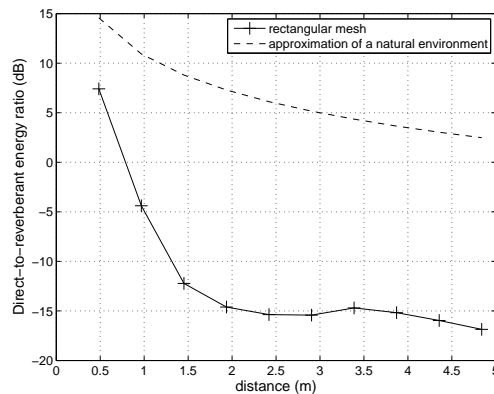


Figure 16: *Direct-to-reverberant energy ratio as a function of distance in the rectangular mesh.*

## References

- [1] P. Zahorik, D.S. Brungart, and A.W. Bronkhorst. Auditory distance perception in humans: A summary of past and present research. *Acta Acustica united with Acustica*, 91:409–420, 2005.
- [2] P. Zahorik. Auditory display of sound source distance. In *Proceedings of the 2002 International Conference on Auditory Display, Kyoto, Japan*, July 2002.
- [3] S.H. Nielsen. *Distance Perception in Hearing*. PhD thesis, Aalborg University, Denmark, 1991.
- [4] D.H. Mershon and L.E. King. Intensity and reverberation as factors in the auditory perception of egocentric distance. *Perception and Psychophysics*, 18:409–415, 1975.
- [5] B. Shinn-Cunningham. Creating three dimensions in virtual auditory displays. In *Proceedings of the 9<sup>th</sup> International Conference on Human-Computer Interaction, New Orleans, Louisiana, USA*, pages 604–608, August 2001.
- [6] M.B. Gardner. Distance estimation of  $0^\circ$  or apparent  $0^\circ$ -oriented speech signals in anechoic space. *J. Acoust. Soc. Am.*, 47(1):47–53, 1969.
- [7] P.D. Coleman. Failure to localize the source distance of an unfamiliar sound. *J. Acoust. Soc. Am.*, 34(3):345–346, March 1962.

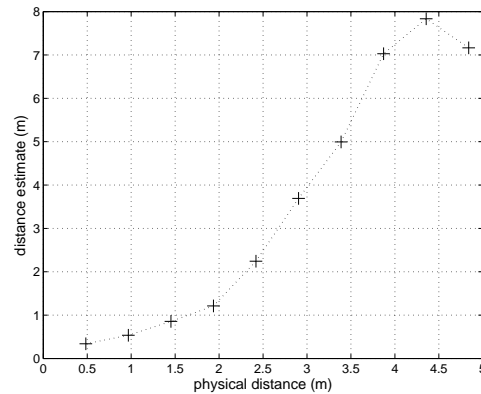


Figure 17: *Distance estimate according to Bronkhorst and Houtgast’s model.*

- [8] N.I. Durlach, A. Rigopoulos, X.D. Pang, W.S. Woods, A. Kulkarni, H.S. Colburn, and E.M. Wenzel. On the externalization of auditory images. *Presence: Teleoperators and Virtual Environments*, 1(2):251–257, 1992.
- [9] A.W. Bronkhorst and T. Houtgast. Auditory distance perception in rooms. *Nature*, 397:517–520, 1999.
- [10] H. Møller, M. Sørensen, D. Hammershøi, and C.B. Jensen. Head-related transfer functions of human subjects. *J. Audio Eng. Soc.*, 43:300–321, 1995.
- [11] T. Houtgast and S. Aoki. Stimulus-onset dominance in the perception of binaural information. *Hearing Research*, 72:29–36, 1994.
- [12] B.C.J. Moore, B.R. Glasberg, C.J. Plack, and A.K. Biswas. The shape of the ear’s temporal window. *J. Acoust. Soc. Am.*, 83(3):1102–1116, 1988.
- [13] F. Fontana and D. Rocchesso. A physics-based approach to the presentation of acoustic depth. In *Proceedings of the 2003 International Conference on Auditory Display, Boston, MA, USA*, pages 79–82, July 2003.
- [14] D. T. Murphy and M. Beeson. The KW-boundary hybrid digital waveguide mesh for room acoustics applications. *IEEE transactions on audio, speech and language processing*, 15(2):552–564, February 2007.
- [15] L. Savioja. *Modeling Techniques for Virtual Acoustics*. PhD thesis, Helsinki University of Technology, 1999.
- [16] J. Huopaniemi, L. Savioja, and M. Karjalainen. Modeling of reflections and air absorption in acoustical spaces – a digital filter design approach. In *IEEE 1997 Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, New York, USA*, October 1997.

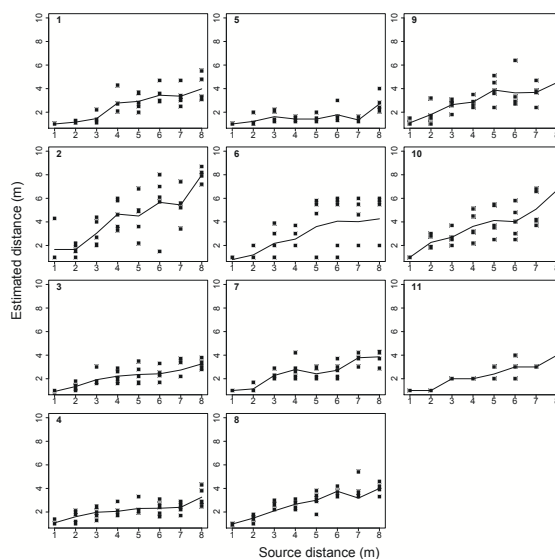


Figure 18: *Exp. III: Individual apparent distance judgments in the rectangular DWM as a function of physical distance. Each plot displays data for one subject. Means are connected by a line.*

- [17] A. Kelloniemi. Frequency-dependent boundary condition for the 3-D digital waveguide mesh. In *Proceedings of the 9<sup>th</sup> Int. Conference on Digital Audio Effects (DAFx-06)*, Montreal, Canada, September 2006.
- [18] F. Fontana. *Physics-based models for the acoustic representation of space in virtual environments*. PhD thesis, University of Verona, Italy, 2003.
- [19] A. Kelloniemi. *Room Acoustics Modeling with the Digital Waveguide Mesh Boundary Structures and Approximation Methods*. PhD thesis, Helsinki University of Technology, Finland, 2006.
- [20] W. Jesteadt, C.C. Wier, and D.M. Green. Intensity discrimination as a function of frequency and sensation level. *J. Acoust. Soc. Am.*, 61(1):169–177, 1977.
- [21] P. Zahorik. Direct-to-reverberant energy ratio sensitivity. *J. Acoust. Soc. Am.*, 112(5):2110–2117, Nov. 2002.
- [22] J.M. Loomis, R.L. Klatzky, J.W. Philbeck, and R.G. Golledge. Assessing auditory distance perception using perceptually directed action. *Perception & Psychophysics*, 60(6):966–980, 1998.
- [23] ITU (International Telecommunication Union). Recommendation BS.1534-1: Method for the subjective assessment of in-



Table 3: Exp. III: Individual fitted linear and power function parameters for each subject.

subject n.	Linear fit			Power fit		
	slope	intercept	$r^2$	$a$	$k$	$r^2$
1	0.45	0.48	0.69	0.48	1.75	0.76
2	0.84	0.53	0.65	0.61	2.11	0.64
3	0.30	0.81	0.63	0.36	1.88	0.67
4	0.24	1.03	0.55	0.28	2.08	0.58
5	0.15	0.87	0.31	0.18	1.96	0.35
6	0.54	0.42	0.40	0.54	1.65	0.43
7	0.41	0.67	0.73	0.44	1.85	0.75
8	0.42	0.77	0.55	0.43	1.91	0.40
9	0.44	1.01	0.56	0.45	2.10	0.62
10	0.70	0.56	0.72	0.57	2.02	0.75
11	0.41	0.46	0.83	0.44	1.76	0.86

intermediate quality levels of coding systems (MUSHRA), 2003.  
<http://www.itu.int/rec/recommendation.asp?type=folders&parent=REC-BS.1534>.

- [24] F. Fontana and D. Rocchesso. Auditory distance perception in the acoustic pipe. *ACM Transactions on Applied Perception*, 5(3):Article 15, 2008, in press.
- [25] E. Vincent. MUSHRAM: a Matlab interface for MUSHRA listening tests, 2005. <http://www.elec.qmul.ac.uk/people/emmanuelv/mushram/>.
- [26] P. Zahorik. Assessing auditory distance perception using virtual acoustics. *J. Acoust. Soc. Am.*, 111(4):1832–1846, 2002.
- [27] P. Zahorik. Challenges in the auditory display of distance information. In *Proceedings of the 19<sup>th</sup> International Congress of Acoustics, Madrid, Spain*, September 2007.
- [28] F. Wightman and D. Kistler. Measurement and validation of human HRTFs for use in hearing research. *Acta Acustica United with Acustica*, 91:429–439, 2005.