RECOGNITION OF DISTANCE CUES FROM A VIRTUAL SPATIALIZATION MODEL

Laura Ottaviani, Federico Fontana and Davide Rocchesso

Dipartimento di Informatica
Università degli Studi di Verona
Strada Le Grazie, 15 - 37134 Verona, Italy
{ottaviani,fontana,rocchesso}@sci.univr.it

ABSTRACT

Emerging issues in the auditory display aim at increasing the usability of interfaces. In this paper we present a virtual resonating environment, which synthesizes distance cues by means of reverberation. We realize a model that recreates the acoustics inside a tube, applying a numerical scheme called Waveguide Mesh, and we present the psychophysical experiments we have conducted for validating the information about distance conveyed by the virtual environment.

1. INTRODUCTION

Research in multi-modal interface design aims at exploiting the multiple sensations of the user, to gain effectiveness in the interaction with the system.

The human auditory system is able to perform not only a recognition of the sound source location and the sound characteristics, but also of the environment where this source is located.

Kubovy and Van Valkenburg define an attribute to be *indispensable* if and only if it is a prerequisite of perceptual numerosity [1]. They state that there are different indispensable attributes for the visual and the hearing channel. For the visual channel, space is an indispensable attribute. For the auditory channel, this role is played by pitch, that becomes an indispensable attribute for hearing numerosity.

However, even if spatial attributes are not as important for hearing as they are for sight, nevertheless range cues become crucial in several situations:

- in auditory warnings, where sounds are used to steer the visual attention;
- to represent events coming from visually occluded or out of sight objects;
- in the design of interfaces for visually impaired users.

The location of a sound source with respect to the listener is defined by three coordinates: azimuth (the angular distance measured along the horizon), elevation (the angular distance measured above the horizon) and distance.

So far the research in this area has focused especially on the directional cues [2] and the stimuli that determine the perceived distance [3]. These stimuli produce cues that are divided in two categories, based on the observer motion state: static and dynamic.

In the experiments we have conducted, listeners perform their tasks from a fixed location in space. Therefore, we analyze only the former category, i.e. *static distance cues*, which play an important role in distance perception when the listener's head is stationary. These cues are [4]:

Intensity which plays a fundamental role, especially with familiar sounds in open space. In the ideal case, intensity in open space decreases of 6 dB for each doubling of the distance between source and listener [5].

Direct-to-reverberant energy ratio that affects distance estimations in environments with sound reflecting surfaces. Reverberant energy is determined mainly by the size of the enclosure and the acoustic properties of its reflecting surfaces. There are also a few studies about outdoor environments that produce reverberation [6].

Spectrum which conveys distance information as well, if the listener has enough familiarity with the original sound. Spectral changes are due to the sound-absorbing properties of the air, for distances greater than 15 m, and to the sound reflection over non-ideal surfaces [7].

Binaural differences that represent an important cue especially for nearby sources [8].

The auditory display aims at increasing the usability of interfaces instead of focusing on the audio quality *per se*. Exaggerating some aspects of the displayed sounds, as in the visual representation field for interfaces [9], contributes to improve usability, as it happens in systems for supernormal auditory localization [10].

There are two approaches that model the effects of the environment characteristics. The *perceptual approach* [11] aims at reproducing the reverberation effects, at the listener's point. This approach provides high-quality rendering, regardless of the physical parameters, given that the psychophysical process that maps the acoustics of a reverberant enclosure is still partially unknown [12]. Moreover, it leads to affordable architectures working in real-time. Nevertheless, most of these realizations do not deal with distance rendering of sound sources.

On the contrary, the *structural approach* aims at modeling environments, focusing on the structural properties that must be rendered, such as the geometry of an enclosure or the materials the wall surfaces are made of. The reverberation effects result as a consequence. Unfortunately, structural models result to be either too resource-consuming or, if the system is simplified to accommodate the hardware requirements, excessively poor in the quality of the audio results.

In this paper we present a virtual resonating environment, aiming at enhancing distance perception by means of reverberation. We model the acoustics inside a tube, using the structural approach. We will introduce the key aspects of the resonator design and then we will focus on its perceptual validation describing and commenting two psychophysical experiments we conducted using this model.

2. THE MODEL OF THE RESONATING ENVIRONMENT

The listening environment we consider is a tubular cavity, with square section and size equal to 9.5×0.45×0.45 m. We set the tube size according to prior investigations, in order to convey an interesting range of distance cues by using a resonating environment that is structurally simple and computationally relatively light. [13] The internal surfaces of the tube are modeled to exhibit natural absorption properties against the incident sound pressure waves. In order to avoid echoes originating from subsequent reflections of the wavefronts along the main direction of wave propagation, the surfaces at the tube ends are modeled to behave as *total* absorbers [14], otherwise reflections along the main axis negatively affect the perception of environmental features.

Although this resonator generates sounds that are quite artificial, due to its geometry, it is able to provide significant distance cues, and a robust rendering of both familiar and unfamiliar sounds.

We modeled the resonator by means of the Waveguide Mesh [15], a particular formulation of the finite difference scheme where wave components traveling along the mesh sum to produce physical (pressure) quantities.

The absorption properties of the boundary were modeled using Digital Waveguide Filters [16], parametrized to damp the low frequency by an absorption coefficient equal to 0.03, progressively increasing with frequency toward 1 [14].

Our model is not provided with a binauralization system. In fact, it is devoted to synthesize cues coming from the external environment. Any kind of range cue that is subject-dependent should be object of investigation for researchers in binaural models [8].

3. THE EXPERIMENTS

We conducted two experiments applying the magnitude estimation method without modulus [17], that is a comparing stimulus to which the experimenter associates a value for reducing the estimated value range. Infact, it has been observed [18, 19] that the module could introduce systematic errors both on the values used by the subjects and on the slope of the computed functions.

We investigated how subjects scaled the perceived distance and, hence, whether our model is effective or not.

The setup involved a PC Pentium III, with a Creative SoundBlaster Live! soundcard. During the first experiment sounds were auditioned through Beyerdynamic DT 770 closed headphones; in the second experiment, the participants sat at a distance of 1.5 m from a pair of Genelec 2029B stereo loudspeakers, 1 m far from each other, and a Genelec subwoofer located in between the loudspeakers.

We will present the two experiments, showing the collected data and comparing the results, in order to evaluate our model in *public* rather than *personal* spaces [20].

3.1. Listening by headphones

Participants. The first experiment involved 12 volunteers (4 female and 8 males), with age between 22 and 40. They study or work at the University of Verona. All of them were naive listeners.

Stimuli. The sound set was synthesized using the following technique. By putting a sound source at one end of the virtual tube, along the main axis, we acquired ten stereophonic impulse responses along positions $x_{10}, \ldots, x_1$, where each one gets closer

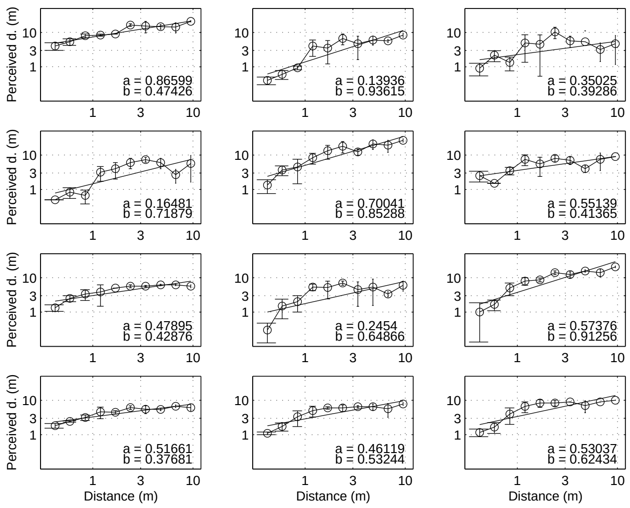

Figure 1: *Headphone listening: Individual distance evaluations together with individual linear regression lines.* a*: intercept.* b*: slope.*

to the sound source by a factor of $\sqrt{2}$, and the first one is equal to 9.5 m. Therefore, the final set X of distances expressed in meters is:

$$X = \{x_i,\, i = 1, \ldots, 10\} = \{0.42, 0.59, 0.84, 1.19, 1.68, 2.37, 3.36, 4.75, 6.71, 9.5\} \quad (1)$$

The right channel of the stereophonic sound accounts for acquisition points exactly standing on the main axis, whereas the left channel accounts for points displaced two junctions far from that axis, this corresponding to an interaural distance of about 15 cm. The impulse responses obtained in this way have been convolved with a short, anechoic sample of a cowbell.

Each stimulus in the set was repeated 3 times in random order, leading to a group of 30 sounds for the experiment.

Procedure. We asked subjects to estimate the perceived distance from the stimuli, using headphones. The participants had to rate each distance with a value in meters (either integer or decimal), starting from the first one, and associating a value to the other ones, proportionally to the first estimation. The experiment was conducted without training. Moreover, we did not set a modulus and, so, the collected values define scales that depend on the individual listeners' judgments. These scales range from 0.2-8 (subject no. 8) to 1-30 (subject no. 5).

The three judgments given for each sound were then geometrically averaged for each subject, and the resulting values were used to calculate a mean average. Subtracting it from the individual averages, we adjusted the listeners' judgments to obtain a common logarithmic reference scaling [19].

Results and observations. In fig. 1 the distance evaluations as functions of the source/listener distance are plotted for each subject, together with the corresponding linear functions obtained by linear regression. The average slope is 0.6093 (standard deviation 0.2062), while the average intercept is 0.4649 (standard deviation 0.2132).

In fig. 2 the perceived distance averaged across values is plotted as function of the source/listener distance, together with the relative regression line ($r^2 = 0.7636$, $F(1, 8) = 25.8385$, $F_{crit}(1, 8)$

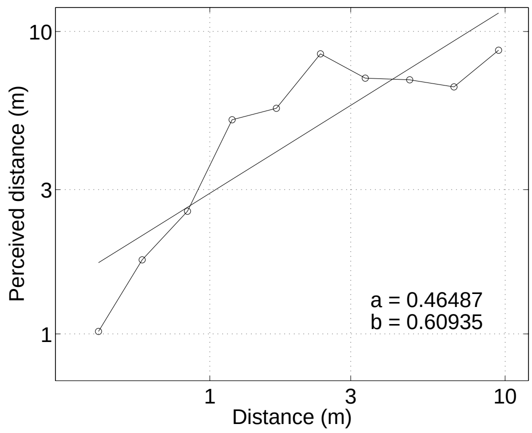

Figure 2: *Headphone listening: Average distance evaluation together with linear regression line. a: intercept. b: slope.*

$= 11.2586$, $\alpha = 0.01$). The r^2 coefficient is significant at $\alpha = 0.01$ and, therefore, the regression line fits well with the subjects' evaluations.

We observed, from this first experiment, that subjects overestimate the distance for sound sources that are close to the listener, and that they reduce this overestimation for greater distances. This result is interesting, since it partially conflicts with other researches [4] that use real sounds for the tests, and report the tendency of listeners to overestimate shorter distances, and underestimate longer distances. In those cases, the point of correct estimation is about $1 - 1.5$ m. In our model, that point is shifted to the far-field.

This result can be interpreted as a consequence of the exaggerated reverberant energy produced by our model.

3.2. Listening by loudspeakers

Participants. The second experiment involved 10 participants (4 females and 6 males), 4 of which participated also to the first experiment. They were volunteers, that work or study in our department and with age between 23 and 32.

Stimuli and Procedure. The set of stimuli was the same as for the first experiment, but the subjects listened to the sounds using loudspeakers. The listeners sat 1.5 m far from the loudspeakers and were blindfolded, in order to minimize the influence of factors external to the experiment. Listeners had to evaluate the distance of the sound source from the listening point communicating its value to the experimenter, who wrote down the data. The first value, as in the previous test, determined the subjective scale

Four participants were involved also in the first experiment.

Results and observations. In fig. 3 we report, for each subject, the distance evaluations as functions of the source/listener distance, together with the corresponding linear functions obtained by linear regression. The average slope is 0.5337 (standard deviation 0.1741), while the average intercept is 0.5034 (standard deviation 0.3573).

In fig. 4 the perceived distance averaged across values is plotted as function of the source/listener distance, together with the relative regression line ($r^2 = 0.8512$, $F(1, 8) = 45.7603$, $F_{crit}(1, 8) = 11.2586$, $\alpha = 0.01$). The r^2 coefficient is significant at $\alpha = 0.01$ and, therefore, the regression line fits well with the subjects' evaluations.

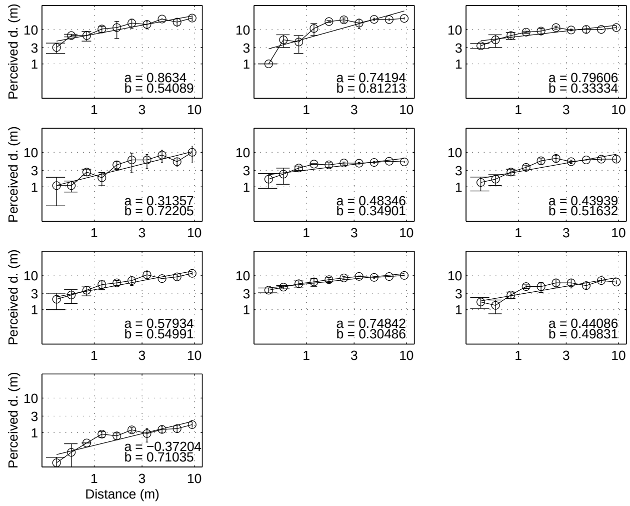

Figure 3: *Loudspeaker listening: Individual distance evaluations together with individual linear regression lines. a: intercept. b: slope.*

We can see that conducting the test using loudspeakers led to results similar to the test with listeners wearing headphones. In fact, in both cases, it is evident that there is a distance overestimation for closer sound sources, that reduces as the distance increases.

There is only one subject (no. 10) whose individual scale ranges between 0.1-2 and who perceived all the sound sources closer than the other listeners. However, during the talk/questionaire after the test, this participant didn't refer to any difficulty in performing the task required.

Furthermore, there is no evident difference between judgements of naive participants, and subjects "trained" by the previous experiment.

4. CONCLUSIONS

A comparison between the two experiments gives interesting hints. First of all, the subjects' responses are similar in both the reproduction conditions.

There is a branch of auditory display that studies the differences existing between headphone and loudspeaker presentation of spatialized sounds [21]. In our model we have not added any specific adaptation to different devices. Nevertheless, our model behaves the same way with both reproduction systems.

Moreover, there is an exaggeration especially in rendering close sound sources, probably due to the amount of reverberant energy existing in that case. The point of correct estimation, in both the reproduction scenarios, is far away from results obtained by other researchers [4]. For this reason, our virtual resonating environment could be adopted in the setup of auditory displays where sounds in the far-field must be presented, without any particular requirement on the reproduction device.

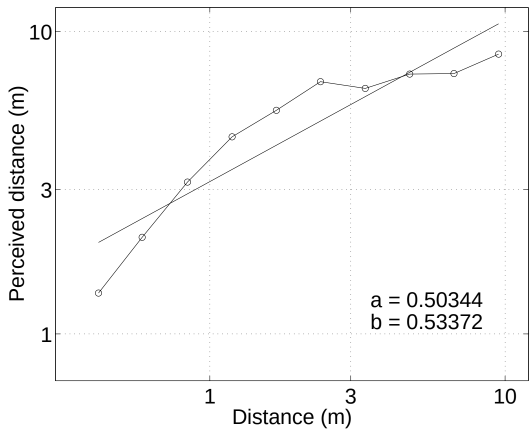

Figure 4: *Loudspeaker listening: Average distance evaluation together with linear regression line. a: intercept. b: slope.*

5. ACKNOWLEDGMENTS

We are grateful to Federico Beghini and Fabio Deboni, who designed preliminary virtual listening environments. We also thank the volunteers who participated in the experiments.

6. REFERENCES

[1] M. Kubovy and D. Van Valkenburg, "Auditory and visual objects", *Cognition*, 80:97–126, 2001.

[2] D. Rocchesso, "Spatial effects", in *DAFX: Digital Audio Effects*, U. Zölzer, Ed., pp. 137–200. John Wiley and Sons, Inc., New York, 2002.

[3] J.M. Loomis and R.L. Klatzky and J.W. Philbeck and R.G. Golledge, "Assessing auditory distance perception using perceptually directed action", *Perception & Psychophysics*, 60(6):966–980, 1998.

[4] P. Zahorik, "Assessing auditory distance perception using virtual acoustics", *Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1832–1846, 2002.

[5] P. M. Morse and K. U. Ingard, *Theoretical Acoustics*, McGraw-Hill, New York, 1968, Reprinted in 1986, Princeton Univ. Press, Princeton, NJ.

[6] D.G. Richards and R.H. Wiley, "Reverberations and amplitude fluctuations in the propagations of sound in a forest: Implications for animal communication", *Am. Nat.*, 115:381–399, 1980.

[7] J. Blauert, *Spatial Hearing: the Psychophysics of Human Sound Localization*, MIT Press, Cambridge, MA, 1983.

[8] D. S. Brungart, N. I. Durlach, and W. M. Rabinowitz, "Auditory localization of nearby sources. ii. localization of a broadband source", *Journal of the Acoustical Society of America*, vol. 106, pp. 1956–1968, 1999.

[9] B. H. Thomas and P. Calder, "Applying cartoon animation techniques to graphical user interfaces", *ACM Trans. on Computer-Human Interaction*, vol. 8, no. 3, pp.198-222, September 2001.

[10] N. I. Durlach, B. G. Shinn-Cunningham, and R. M. Held, "Supernormal auditory localization. i. general background", *Presence: Teleoperators and Virtual Environments*, vol. 2, no. 2, pp. 89–103, 1993.

[11] W. G. Gardner, "Reverberation algorithms", in *Applications of Digital Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds., pp. 85–131. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.

[12] L. L. Beranek, "Concert hall acoustics - 1992", *Journal of the Acoustical Society of America*, vol. 92, no. 1, pp. 1–39, July 1992.

[13] F. Fontana, D. Rocchesso and L. Ottaviani, "A Structural Approach to Distance Rendering in Personal Auditory Displays", International Conference on Multimodal Interfaces, ICMI'02, Pittsburgh, USA, October 14-16,2002.

[14] H. Kuttruff, *Room Acoustics*, Elsevier Science, Essex, England, 1991, Third Ed.; First Ed. 1973.

[15] S. A. Van Duyne and J. O. Smith, "Physical modeling with the 2-D digital waveguide mesh", in *Proc. Int. Computer Music Conf.*, Tokyo, Japan, 1993, ICMA, pp. 40–47.

[16] S. A. Van Duyne and J. O. Smith, "A simplified approach to modeling dispersion caused by stiffness in strings and plates", in *Proc. Int. Computer Music Conf.*, Aarhus, Denmark, Sept. 1994, ICMA, pp. 407–410.

[17] S. S. Stevens, "The direct estimation of sensory magnitude-loudness", *American Journal of Psychology*, vol. 69, pp. 1–25, 1956.

[18] F. Nowell-Jones, "Overview of psychophysical scaling methods" in *Handbook of perception*, E. C. Carterette and M. P. Friedman (ed.), vol.II, pp.343-360, Academic Press, New York, 1974.

[19] T. Engen, "Psychophysics. I. Discrimination and detection. II Scaling" in *Woodworth & Schlosberg's Experimental Psychology*, J. K. kling and L. A. Riggs (ed.), 3rd ed., pp.11-86, Methuen, London, 1971.

[20] W. L. Martens, "Psychophysical calibration for controlling the range of a virtual sound source: Multidimensional complexity in spatial auditory display", in *Proc. International Conference on Auditory Display*, Espoo - Finland, Aug. 2001, pp. 197–207.

[21] J.-M. Jot, "Real-time spatial processing of sounds for music, multimedia and interactive human-computer interfaces", *Multimedia systems*, 7, pp.55-69, 1999.