SUBJECTIVE EVALUATION OF AUDIO EGOCENTRIC DISTANCE IN REAL AND VIRTUAL ENVIRONMENTS USING WAVE FIELD SYNTHESIS.

S Moulin Orange Labs, Lannion, FRANCE
R Nicol Orange Labs, Lannion, FRANCE
L Gros Orange Labs, Lannion, FRANCE
P Mamassian Laboratoire Psychologie de la Perception, Paris, FRANCE

1 INTRODUCTION

Due to the recent generalization of 3D visual technologies, anybody can experience 3D audio-visual: stereoscopic movies, 3DTVs, smartphones, or even virtual reality are various examples of 3D application. Most of the time, this experience results in the addition of 3D visual plus 2D audio technologies like stereophony or 5.1 surround systems for instance. Does this new visual dimension need the introduction of a new appropriate audio technology? Maybe the use of 3D audio technologies such as binaural[1,2], holophonic[3,4,5], or multichannel reproduction system could enhance our audio-visual experience in terms of spatial consistency or naturalness.

Up to now, the field of audio-visual perception was poorly investigated especially in the distance dimension. For instance, audio-visual phenomena such as ventriloquist effect[6] or spatial integration window[7] are quite well understood for the directional localization but there are only few studies about these effects in the distance dimension. Before dealing with these cross-modal issues, we need to study the unimodal distance perception to understand better the underlying perceptual properties. As a first step, this paper focuses on auditory perception and intent to answer this question: to which extent a given 3D audio technology is able to render the auditory distance?

This paper describes experiments about auditory distance perception of real and virtual sound sources. The virtual sound environment is provided using Wave Field Synthesis[3]. First, we propose a brief recall of auditory distance perception mechanisms. Measurement protocols for distance estimation and results of previous studies about real and virtual sound sources are also presented. Then, our experimental setup is described. The perceived distances are finally analysed as a function of physical or rendered distances depending on the experiment.

2 AUDITORY DISTANCE PERCEPTION

2.1 Acoustic and Non-Acoustic Distance Cues

Scientific community agrees that auditory distance perception of a static sound source by a static listener mainly depends on four acoustics factors[8,9,10,11]. First, sound level and direct-to-reverberant energy ratio are probably the better-known and most impacting acoustic distance cues[12]. Sound level acts as a relative cue whereas reverberation is an absolute cue. In an anechoic environment, the sound level loss is 6 dB per doubling distance. This loss is different in an echoic room since the direct-to-reverberant ratio changes as a function of the distance of sound sources due to the contribution of wall reflections. In addition, when a sound source moves away from a listener, its spectral properties are also modified. Due to the air absorption, high frequencies are progressively attenuated with the distance but this effect is potentially influent for distance farther than 15 meters[9]. Finally, binaural differences can potentially be useful to estimate distance of near-field sound sources (within approximately 1 meter)[13,14].

In addition to these acoustic cues, some non-acoustic cues can be helpful to estimate the distance of sound sources in static situation. Vision can obviously give important information about an object location but can also bias the auditory perception. This bias is known as the "ventriloquist effect" for directional localization[6] but some studies mentioned similar effect for the distance localization[15,16]. Other studies have shown that distance perception can also be influenced by the familiarity with the sound stimulus. For instance, it was suggested that distance judgment accuracy can be improved by the use of a familiar stimulus such as speech[17], or with past experience[18,19].

2.2 Distance Measurement Protocols

There are many different distance measurement protocols but they can be classified in four main categories[20]: verbal reports, perceptually guided actions, imagined actions, and perceptual matching protocols. Verbal estimation protocols can consist in a judgment of distance using scales with familiar units (meters, feet). This category also includes direct report method (without scales)[21]. In perceptually guided protocols, listeners have to perform an action to estimate the distance of the objects. For instance, they can have to walk directly to a target[10] ("direct walking" protocol) or following an indirect path[22] ("triangulated blind walking" protocol). Imagined actions include protocols in which participants have to imagine the needed time of walk or number of steps to reach a target for instance. In perceptual matching protocols, subjects estimate the distance of a target object by moving a matching sound object[23]. These protocols include "perceptual bisection" or "method of adjustment" and result in a comparison of different perceptual information.

Loomis et al.[10] found that verbal reports from observers are in agreement with distance estimations obtained using both direct and indirect walking methods even if verbal measures show greater variability. All these methodologies of distance estimation have pros and cons but it is important to notice that verbal reports don't imply other modalities to estimate distances. Indeed, perceptually guided actions induce visual and/or motor responses which have their own variability. In addition, experimental measures from imagined actions need to be converted into distances.

2.3 Distance Perception of Sound Sources

The distance perception of sound sources hasn't been as investigated as localization of sound sources in azimuth. Nevertheless, Zahorik made a review[11] of more than 80 experiments about this subject. There is a large variability between compared experimental conditions: environments, stimuli, azimuth and distance range can be different. Despite these experimental differences, most results indicate that participants tend to underestimate the physical sound source distances. Zahorik affirmed that the relation between perceived distances and physical distances is well approximated by compressive power function of the form $d_p = k(d_r)^a$ where d_p is the perceived distance, d_r is physical distance, a is the power function exponent, and k is a constant. He found that the average value of "k" is 1.32, and the exponent of power functions "a" is less than one ($\bar{a} = 0.54$). He also noticed an overestimation of distances for near sound objects ($d_r < 1.5$ m).

Different experiments were recently conducted about distance perception specifically in virtual environments. These studies cover most of 3D sound technologies such as binaural[21,24,25], or holophonic rendering for static[26,27] or moving subjects[22,28]. They show that auditory performances in virtual and real environments are about the same in terms of distance perception. Auditory distances are generally underestimated and are consistent with Zahorik's compressive model.

3 EXPERIMENTAL SETUP

We conducted two experiments about egocentric auditory distance perception of real sound sources (experiment 1), and virtual sound sources using Wave Field Synthesis (experiment 2). It should be highlighted that this work doesn't attempt to compare directly real vs. virtual sound sources distance estimations. The aim of the first experiment is to validate our protocol of distance measurement, especially to ensure that the selected reporting method is relevant. The second experiment aims at estimating to which extent the audio depth can be simulated by a WFS system.

3.1 Environment

Experiments take place in a damped room in Orange Labs. In this room, the reverberation time is about 350 ms (T_{60}) and the background noise level is less than 30 dB(A). An acoustically transparent curtain is placed in front of the participant in order to prevent him to see the sound sources or to get any idea of the room size (about 20 m^2).

The audio rendering is provided using a Motu 24I/O interface controlled by a Motu 424 PCIe sound card. The Wave Field Synthesis implementation and signal processing is made using Max/MSP.

3.1.1 Presentation of real sound sources

In the first experiment, eight Studer loudspeakers are located at 1, 1.5, 2, 2.5, 3, 3.5, 4, and 5 meters from the listening position. As it can be seen in Figure 1, loudspeakers are slightly shifted in elevation and azimuth in order to reduce the direct wave diffraction. These shifts are restricted to a spatial window of ±1.5° in azimuth and ±2.5° in e levation based on localization blur[9]. An equalization procedure is performed to match the spectral responses of loudspeakers. To do so, a microphone is placed at the listening position and the impulse response of each loudspeaker is measured for its actual location. The equalization process is calculated only taking into account the direct wave in order to keep the "natural" room response at the different distances. Loudspeakers are calibrated to 68 dB(A) at one meter from their position.

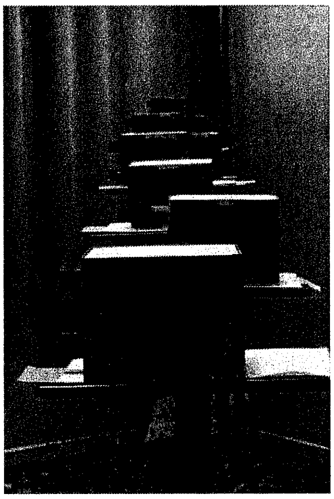

Figure 1: real sound sources placed in front of participants.

3.1.2 Presentation of virtual sound sources

In the second experiment, virtual sound sources are simulated by WFS using an array of 24 Studer loudspeakers spaced by 9 cm. The WFS array is placed 3 meters away from the listener (see Figure 2).

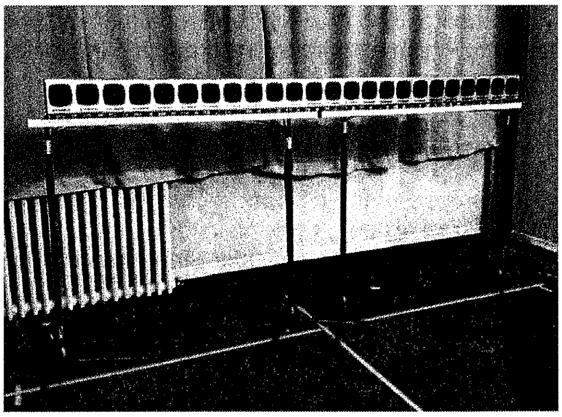

Figure 2: Wave Field Synthesis array.

Loudspeakers spectral responses are equalized using impulse responses measured at the listening position. The equalization is implemented so that all the individual contributions of each loudspeaker are identical at the listening point. The resulting acoustic field is as it would be if loudspeakers were ideal omnidirectional sound sources (as specified by WFS theory[3]). The simulated source corresponds to a monopole emitting 73 dB(A) at one meter.

3.1.3 Auditory stimuli

Three stimuli are tested in order to cover a wide range of familiarity and spectral cues: female speech, ringtone, and white noise bursts. Two of these stimuli are commonly used for localization tasks but for different reasons. Speech signal is frequently used because it is a familiar sound whereas white noise is used for its spectral properties. The third stimulus of this study (ringtone) is chosen because of its potential strong semantic relation with a visual object. Indeed, future studies will involve audio-visual object distance estimation.

3.2 Test Procedure

Listeners are static and blindfolded during all the test duration. In the first experiment, participants are asked to estimate eight distance conditions between 1 and 5 meters. The second experiment involves two additional distances (at 7 and 10 meters) for a total number of ten distance conditions. In both experiments, distance estimations are made using a direct report method. For each distance condition, participants have to report their estimation on a keyboard with 0.1 m accuracy.

Every distance condition is repeated five times. It results in a total of 40 conditions per stimulus in experiment 1 and 50 conditions per stimulus in experiment 2. The presentation order of stimuli changes every four spectators in order to present each possible combination to the same number of participants. The distance conditions are randomly varied for each participant but keep unchanged for the three stimuli. The average test duration is 30 minutes for experiment 1 and about 40 minutes for experiment 2.

3.3 Panel Composition

The panel consists of 24 participants in both experiments (11 women and 13 men for the first experiment and 13 women and 11 men for the second) whose average ages are 31 and 30 years. All participants have experience in listening tests, but none of them took part in a subjective test which involves auditory distance estimations.

4 RESULTS

4.1 Experiment 1

A variance analysis (ANOVA) is performed on participant judgments considering two between-group factors: "Stimuli" (three levels), and "Distances" (eight levels). Figure 3 represents mean distance estimations and associated 95% confidence intervals for 22 participants. Indeed, two subjects were excluded from this analysis because of high variability in comparison to panel results (more than five times the standard deviation); especially for the nearest loudspeaker for which the inter-individual variability is the lowest.

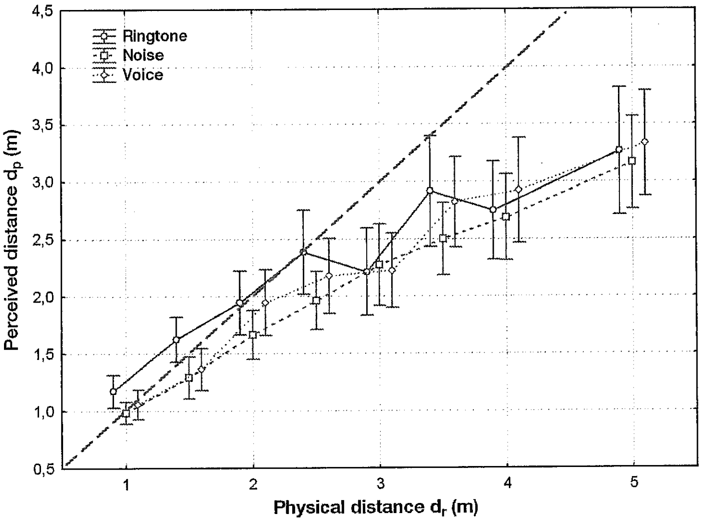

Figure 3: means and 95% CI of perceived distances of real sound sources placed in the distance range [1 - 5] meters.

As expected, results of the first experiment show that participants underestimate auditory distances and are more accurate for nearer loudspeakers. The ANOVA reveals that participants are able to discriminate distances (F(7,147)=95.79, p<0.001) but there is no significant effect of the stimulus (F(2,42)=2.956, p>0.05). Table 1 presents the calculated constant "k" and exponent "a" based on Zahorik's model described by the equation $d_p = k(d_r)^a$ (see part 2.4).

Stimulus	k	a	R^2
Ringtone	1.237	0.62	0.958
Noise	0.986	0.74	0.997
Voice	1.075	0.73	0.979

Table 1: compressive power function values and associated coefficient of determination (R^2) for the three stimuli in the case of real sound sources.

Even if values in Table 1 slightly vary according to stimuli, results found in experiment 1 are in a good agreement with previous studies. Moreover, high values of R^2 factor indicate that predictions by Zahorik's compressive model perfectly match the experimental results. According to all these observations, the employed protocol measurement can reasonably be validated to estimate distance. Thus, it will be used for further studies, starting with experiment 2 about auditory distance perception of virtual sound sources.

4.2 Experiment 2

In the second experiment, participants have to estimate the distance of virtual sound sources. An ANOVA is also performed on participant judgments but in this case the "Distances" between-group factor has ten levels. Figure 4 represents mean distance estimations and associated 95% confidence intervals for the 24 participants.

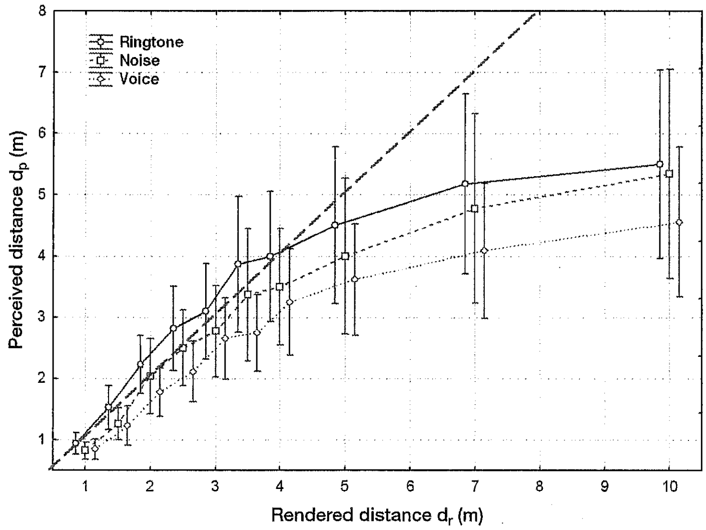

Figure 4: means and 95% CI of perceived distances of virtual sound sources placed in the distance range [1 - 10] meters.

First, confidence intervals presented on Figure 4 indicate that participant judgements are more accurate for nearer virtual sound sources. This plot also shows that differences between stimuli increase with the rendered distance. "Voice" stimulus is generally perceived closer than "Ringtone" or "Noise" with a smaller variability (for 5, 7, and 10 meters conditions). The ANOVA confirms this effect of stimulus (F(2,46)=6.126, p<0.005). Previous studies[29,30] have shown that distance perception of speech could be influenced by the speech enunciation and have found that a conversational speech is naturally perceived closer than a shouted speech. In the case of our experiment, the voice stimulus is comparable to a conversational speech. Thus, there might be an additional underestimation of distance for farther conditions as a speaker would tend to shout if the listener was located a few meters away from him in a natural environment. This could explain a greater compression of the perceived distance of "Voice" stimulus. As for the variability of estimations, Coleman[17] found that using a familiar stimulus can increase distance judgement accuracy which is consistent with our results.

The most important result is that participants are able to discriminate distance conditions in the case of virtual sound sources (F(9,207)=36.143, p<0.001). Table 2 presents the calculated "k" and "a" values of compressive functions explained in part 2.4.

Stimulus	k	a	R^2
Ringtone	1.21	0.78	0.919
Noise	1.03	0.82	0.935
Voice	1.00	0.75	0.945

Table 2: compressive power function values and associated coefficient of determination (R^2) for the three stimuli in the case of virtual sound sources.

Table 2 indicates that compressive power functions suit quite well to experimental data because of high values of R^2 factors ($R^2 > 0.91$). Nevertheless, a logarithmic model of the form $d_p = A. \ln(d_r) + B$ could be more appropriate to these results. Indeed, in this case R^2 values would be greater than 0.98 for all stimuli.

5 CONCLUSION

In this paper, two experiments about egocentric distance perception of real and virtual sound sources are described. The first experiment with real sources validates the measurement protocol. Results show a good agreement with previous studies. The second experiment aims at estimating to which extent Wave Field Synthesis can create auditory events at different distances. The perceived distance of 10 virtual sound sources placed in front of and behind a WFS array (between 1 and 10 meters) is measured. Results show that perceived distances are more or less compressed according to the stimulus but participants are able to discriminate the different distance conditions.

Further work will explore the egocentric visual distance perception of virtual objects provided by a stereoscopic projector, using the same measurement protocol. Then, experiments will be conducted in the case of cross-modal condition in terms of egocentric distance estimations and audio-visual integration window in distance dimension.

6 AKNOWLEDGMENTS

The authors would like to thank Olivier Warusfel and Thibaut Carpentier for their technical support on Wave Field Synthesis implementation.

7 REFERENCES

1. H. Møller., Fundamentals of binaural technology, Applied Acoustics, Vol. 36, 171-218. (1992).
2. R. Nicol., Binaural technology, AES Monograph. (2010).
3. A.J. Berkhout, D. de Vries, and P. Vogel., Acoustic control by Wave Field Synthesis, J.Acoust.Soc.Am, 93(5), 2764–2778. (1993).
4. M.A. Gerzon., Periphony: With-eight sound reproduction, J.Audio.Eng.Soc., 21(1), 2-10. (1973).
5. J. Daniel, R. Nicol, and S. Moreau., Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging, Proc. 114th AES Convention, Amsterdam. (2003).
6. I.P. Howard, and W.B. Templeton., Human spatial orientation, Wiley, New York. (1966).
7. J. Lewald, W.H. Ehrenstein, and R. Guski., Spatio-temporal constraints for auditory-visual integration, Behavioural Brain Research, 121(1-2), 69-79. (2001).
8. D.H. Mershon, and L.E. King., Intensity and reverberation as factors in the auditory perception of egocentric distance, Percept. Psychophys, Vol. 18, 409-415. (1975).
9. J. Blauert., Spatial Hearing, The Psychophysics of Human Sound Localization, MIT Press, Cambridge. (1997).
10. J.M. Loomis, R.L. Klatzky, J.W Philbeck, and R.G. Golledge., Assessing auditory distance perception using perceptually directed action, Percept. Psychophys, Vol. 60, 966-980. (1998).
11. 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, Vol. 91, 409-420. (2005).
12. B.G. Shinn-Cunnigham., Distance Cues for Virtual Auditory Space, Proc. IEEE Pacific-Rim Conf. on Multimedia, 227-230. (2000).

13. B.G. Shinn-Cunnigham, N. Kopco, and T. Martin., Localizing nearby sound sources in a classroom: Binaural room impulse responses, J.Acoust.Soc.Am, 117(5), 3100-3115. (2005).
14. D.S. Brungart, and W.M. Rabinowitz., Auditory localization of nearby sources I: head-related transfer functions, J.Acoust.Soc.Am, 106(3), 1465–1479. (1999).
15. M.B. Gardner., Proximity image effect in sound localization, J.Acoust.Soc.Am, 43(1), 163-163. (1968).
16. D.H. Mershon, D.H. Desaulniers, T.L.J. Amerson, and S.A. Kiefer., Visual capture in auditory distance perception: Proximity image effect reconsidered, J. Aud. Res., 20(2), 129-136. (1980).
17. P.D. Coleman., Failure to localize the source distance of an unfamiliar sound, J.Acoust.Soc.Am, 34(3), 345-346. (1962).
18. N. Kopco, M. Schoolmaster, and B.G. Shinn-Cunningham., Learning to judge distance of nearby sounds in reverberant and anechoic environments, Proc. Joint CFA/DAGA'04, Strasbourg. (2004).
19. B.G. Shinn-Cunningham., Learning reverberation: considerations for spatial auditory displays, Proc. ICAD, 126-134, Atlanta. (2000).
20. E. Klein, J.E. Swan, G.S. Schmidt, M.A. Linvingston, and O.G. Staadt., Measurement protocols for medium-field distance perception in large-screen immersive displays, PRO. IEEE Virtual Reality, 107-113. (2009).
21. N. Côté, V. Khoel, M. Paquier, and F. Devillers., Interaction between auditory and visual distance cues in virtual reality applications, Proc. Forum Acusticum 2011, 1275–1280, Aalborg. (2011).
22. M. Rébillat, X. Boutillon, E. Corteel, and F.G.B. Katz., Audio, visual, and audio-visual egocentric distance perception by moving subjects in virtual environments, ACM Trans. Appl. Percept., 9(4), 19:1-19:17. (2012).
23. H. Wittek, S. Kreber, F. Rumsey, and G. Theile., Spatial perception in wave field synthesis rendered sound fields: distance of real and virtual nearby sources, Pres. at 116th AES Convention, Berlin. (2004).
24. P. Zahorik., Assessing auditory distance perception using virtual acoustics, J.Acoust.Soc.Am, 111(4), 1832-1846. (2002).
25. P.W. Anderson, and P. Zahorik., Auditory and visual distance estimation, Proc. Meetings on Acoustics, Vol 12, 050004. (2011).
26. G. Kearney, M. Gorzel, H. Rice, and F. Boland., Distance perception in interactive virtual acoustic environments using first and higher order ambisonic sound fields, Acta Acustica with Acustica, 98(1), 61-71. (2012).
27. S. Komiyama, A. Morita, K. Kuorzumi, and K. Nakabayaski., Distance control system for a sound image, Proc. 9th AES International Conference on television sound today and tomorrow. (1991).
28. E. Corteel., Caractérisation et extensions de la wave field synthesis en conditions réelles. Ph.D. dissertation, Université de Paris 6. (2004).
29. M.B. Gardner., Distance estimation of 0° or app arent 0°-oriented speech signals in anechoic space, J.Acoust.Soc.Am, 45(1), 47-53. (1969).
30. D.S. Brungart, and K.R.Scott., The effects of production and presentation level on the auditory distance perception of speech, J.Acoust.Soc.Am, 110(1), 425-440. (2001).