ISSN 1063-7710, Acoustical Physics, 2012, Vol. 58, No. 5, pp. 610–617. © Pleiades Publishing, Ltd., 2012.

ACOUSTIC SIGNAL PROCESSING AND COMPUTER SIMULATION

Material Sound Source Localization through Headphones[1]

Larisa Dunai, Guillermo Peris-Fajarnes, Ismael Lengua Lengua, and Ignacio Tortajada Montaña

Research Center in Graphic Technology, Universitat Politècnica de València, Camino de Vera s/n, 8L, València, 46022 Spain

e-mail: ladu@upv.es

Received May 18, 2011

Abstract—In the present paper a study of sound localization is carried out, considering two different sounds emitted from different hit materials (wood and bongo) as well as a Delta sound. The motivation of this research is to study how humans localize sounds coming from different materials, with the purpose of a future implementation of the acoustic sounds with better localization features in navigation aid systems or training audio-games suited for blind people. Wood and bongo sounds are recorded after hitting two objects made of these materials. Afterwards, they are analysed and processed. On the other hand, the Delta sound (click) is generated by using the Adobe Audition software, considering a frequency of 44.1 kHz. All sounds are analysed and convolved with previously measured non-individual Head-Related Transfer Functions both for an anechoic environment and for an environment with reverberation.

The First Choice method is used in this experiment. Subjects are asked to localize the source position of the sound listened through the headphones, by using a graphic user interface. The analyses of the recorded data reveal that no significant differences are obtained either when considering the nature of the sounds (wood, bongo, Delta) or their environmental context (with or without reverberation). The localization accuracies for the anechoic sounds are: wood 90.19%, bongo 92.96% and Delta sound 89.59%, whereas for the sounds with reverberation the results are: wood 90.59%, bongo 92.63% and Delta sound 90.91%. According to these data, we can conclude that even when considering the reverberation effect, the localization accuracy does not significantly increase.

Keywords: material sounds, Delta sound, sound localization

DOI: 10.1134/S1063771012050077

INTRODUCTION

One of the most important factors for human survival is the ability of sound source localization in real environments. This ability helps humans to avoid obstacles that appear in their way, to prevent dangers, to perceive the differences between different types of sounds as well as to judge their representation, to perceive speech, etc. It is important to remark that the sound source localization in the environment is a difficult task, because of the influence of different acoustical cues, such as reverberation and echo, as well as the huge amount of different noises that are present in this atmosphere.

In this regard, it is important to note that a wide research has been developed on sound source localization [1]. The influence of binaural cues, such as Interaural Time Differences and Interaural Level Differences, on sound source localization has drawn special attention [2, 3].

Also the auditory saltation or the audible threshold has been deeply studied, since they are important factors of sound localization [4–7]. Due to the improvement and great demand of acoustical systems and virtual technologies, sound source localization has been also analysed by using generated or measured Head-Related Transfer Functions (HRTF), by means of the creation of virtual environments [8, 9].

With regards to generated HRTFs, the main task is the computer model used for the creation of the environment. The main problems of the computer algorithm for architectural acoustic rely on the acoustic characteristics of the room, as well as on how to estimate reverberation time or sound quality [10, 11].

It is a great dilemma how human auditory system processes and perceives material sounds. [12, 13], investigated the material perception and the influence of variables that govern the synthesis of material sounds. Actually, it is easy to extract acoustic information from the sounds, but the question is how the human brain captures this information and allows us to classify and localize material sound sources. It is important to remark the influence of material physical properties on material sound perception. Among them, the material elasticity, the volume of the object, the force with which the object has been impacted or the frequency are especially important. Up to date, most of the works have intended to explain how humans perceive two different material sounds [12, 14]. However, no works have been found on localization of azimuth and distance of material sounds.

[1] The article is published in the original.

Material sound perception and localization provide very useful information both for representing the surrounding environment and for virtual reality applications. For example, when an impaired person walks on the street, he combines his hearing abilities and his white cane, in order to detect the obstacles in his path and to avoid them. Also, it is crucial for them to perceive the differences between material sounds. This property helps them to analyse the danger of the obstacles and the environment. The sounds or noises generated by cars, street lights, people walking, opening and closing doors, etc., are important survival cues for visually impaired people. On the other hand, with the development of new acoustic navigation aids for blind people, the selection of the correct acoustic sounds becomes a primordial task. This is due to the fact that the impaired user uses the device during long time periods and, therefore, the implemented sounds must be clear, short, and not interfere with the surrounding noises in order to avoid irritating him. In conclusion, the user safety strongly depends on the selected sounds.

With all these facts in mind, the current study aims to analyse the sound source localization, considering three different sounds: wood, bongo and a Delta sound generated using HRTFs. The study considers two different situations: an anechoic environment (i.e. no reverberation) and an environment with reverberation. The paper explains the simple psychoacoustic experiment developed, which implies the use of a graphic user interface as well as of previously measured HRTF. The reason for selecting wood, bongo and Delta (or click) sounds in this psychoacoustic experiment is motivated by their widely spread presence in the human everyday life; these sounds are very common to the listeners, being very their discrimination. Also, it is important for the psychoacoustic experiments to know the effects of the environmental conditions, such as the reverberation (the importance of which has been pointed out by many authors [15]), which enhances the realism of the sounds, providing a more realistic sensation. The paper analyses the responses obtained in all the categories.

Table 1. Time duration of the sounds

Sound	Time in seconds	
	Anechoic	Reverberation
Wood	0.078	0.548
Bongo	0.104	0.548
Delta (click)	0.078	0.548

PROCEDURE

Sound Sample

Two different sounds are firstly obtained by hitting everyday life objects made of wood and bongo. Besides, a Delta sound (a click) is generated with the Adobe Audition software. Then, each sound is simplified based on additive synthesis techniques that resample these recorded sounds at frequency of 44.1 kHz and sound level of 72 dB. All sounds are recorded and generated in two different environments: anechoic and with reverberation. Each signal has different spectral content and duration (See Table 1). A reverberation effect is applied to each sound by using the CATT-Acoustic software with FireReverb™ Suite by PureVerb™. The environment is created considering the following room dimensions: volume $V = 16$ m^3 (width × length × height = 2 × 4 × 2 m^3), and the following values for the program parameters: absorbing element $A = 0.25$ s, reverberation time $T = 1.4$ s and directivity index (which controls the direct sound to reverberation ratio) $D = 1.0$ dB. Finally, each sound is convolved with a non-individual Head-Related Transfer Function (See Fig. 1). The HRTF is generated using a KEMAR dummy head. It is measured for each spatial position in a frontal plane, both in distance and in azimuth:

$$y(n) = x(n)*h(n), \tag{1}$$

where $x(n)$ represents a sound with n samples and $h(n)$ represents the Head-Related Transfer Function of n samples.

In the convolution process between the HRTFs and the sound, it can happen that the convolution is obtained only for one sound channel. Applying this method, the sounds are converted into monaural

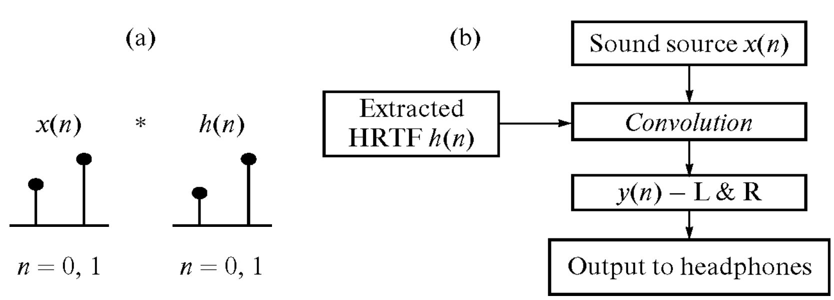

Fig. 1. A convolution representation of the two sequences. In (a), it is shown the mathematical representation of the convolution, in (b) the convolution algorithm between the HRTF and the previously generated signal is presented.

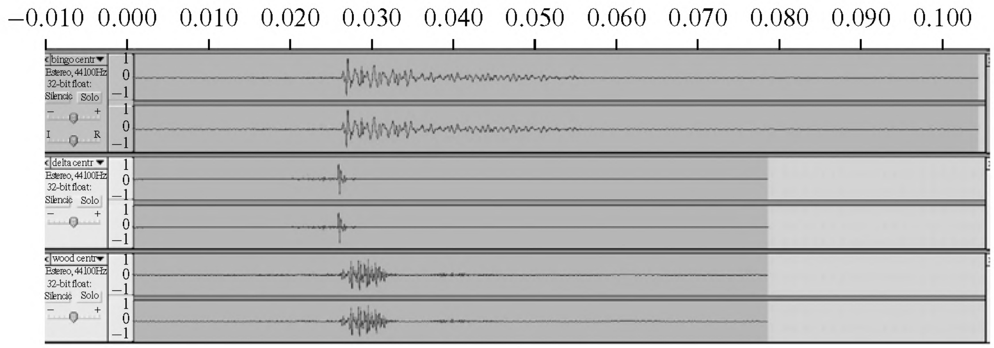

Fig. 2. Representation of sound waves obtained after the convolution between the bongo, Delta sound (click) and wood sounds with the HRTF, for a distance of 0.5 meters and direction of 0°, using the Discrete Fourier Transform method. The x-axis represents the time in seconds (the bongo sound lasts 0.104 s, the delta and wood sounds last 0.078 s), while the y-axis represents the wave amplitude after convolution and normalization (between –1 and 1 values). Each of the three sounds is composed by the two channels (left and right) of the stereo sound.

sounds. Since the objective of this experiment is to create binaural sounds, (1) cannot be used alone. The solution is to apply the Discrete Fourier Transform (DFT) (2), which allows us to transform the signal from time domain to the frequency domain.

$$x(k) = \sum_{n=0}^{N-1} x(n) e^{\frac{-jnk2\pi}{N}}, \qquad (2)$$

where $x(k)$ represents the signal resulting from the DFT and $x(n)$ is the signal convolved. By using DFT (2) we obtain a sequence of complex numbers with length $N/2 + 1$; the output signal contains two channels. Both resulting signals will contain the signal magnitude. Applying the convolution with the HRTFs, the sound is spatialized to appear as if it was located at the corresponding real-world location. This method is used to obtain 3D-stereo sounds to be employed in experiments using headphones. In this way, the sounds listened through the headphones appear as if they were inside the head. In order to represent sounds as if they were coming from the environment, it is important to convolve them with previously generated HRTFs, by using a proper method. In this paper, the convolution has been performed, by using Discrete Fourier Transform.

In order to generate the HRTFs, the Maximum Length Binary Sequences MLBS is applied and later a cross correlation method is used between the system answer and MLBS input. Each sound is processed using Matlab and saved as a Microsoft ".wav" file sampled at 44.1 kHz and with a 16-bit resolution (Fig. 2).

Listeners

Ten young volunteers take part in the experiment. The average age is 25 years. All of them report having normal hearing. Listeners are previously tested with Békésy technique [16, 17] using pure tones between 250 and 8000 Hz. The threshold is within 15 dB in both ears. The technique consists of recording the auditory threshold on an auditory blank. The technique is based on two motors: one motor increases the sound frequency slowly, while the second one causes the tone to become gradually louder. When the listener hears the sound, he pushes the button. In this case, the sound becomes gradually fainter. When he does not hear any sound, he releases the button and the sound gets gradually louder. During the test the frequency is increasing slowly from a very low frequency. All of them are experienced in sound localization tasks.

Method

Stimuli are presented to the users through SENNHEISER HD201 headphones, which are connected to an ACER laptop. For this purpose, an experimental program developed in Flash 8.0 and running on Windows XP is used. The program enables the reproduction of qualitative sounds. The subjects interacted with the experiment using the graphic user interface depicted in Fig. 3.

This experiment is conducted in a conventional laboratory, in which external noises are presented (human speech, noises coming from the street, equipment noises, etc.). The experimental interface consists of a representation of an horizontal plane, corresponding to distances between 0.3 and 5 m and azimuths between 75° (left) and 75° (right). It is divided into 17 areas in distance and in 27 areas in azimuth, resembling a grid. In the top of the interface, sound location data, such as distance, azimuth and mean are specified. At the left side of the interface, the subject can find the 'Repeat the Sound' button. This button allows the subject to repeat the previously listened sound in case he was unable to hear it due to any reason. Below the 'Repeat the Sound' button, a counter of remaining sounds is placed.

The experiment is distributed in two groups: the anechoic material sounds and material sounds with reverberation. Each group consists of three sounds:

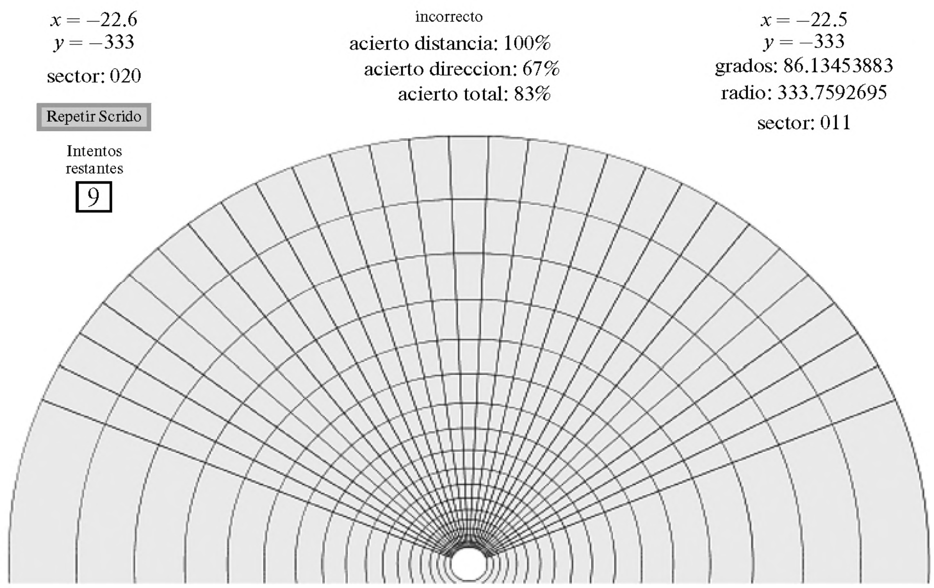

Fig. 3. Graphic user interface, where the upper side of the interface represents the coordinates of the mouse, the results are in percent and the coordinates of the selected sound and the number of rested sounds that are to be reproduced.

wood, bongo, and Delta sound. In each trial, the listener hears a train of sounds in a spatial position randomly delivered by the program and estimates its position, without trying to guess the material. After the user has answered, the next train of sounds is delivered. Each trial consists of ten sounds reproduced in different positions. The method uses the forced choice approach, this is, a sound is presented to the user in a certain location and the listener is required to report whether the sound has appeared at the left, right, center, near or far by pressing one cell of the user interface. Even if he did not understand or listen the sound, he is forced to answer. In other words, 'I do not know' is not an acceptable response. In the experiment, the duration of the sound is not taken into account. Participants can repeat the sound as many times as they decide by pressing the REPEAT button. Then, the listener responds by clicking with the mouse on one of the cells on the interface. The volume control is constant for all subjects. There is no feedback. A typical run lasts 5–8 min, depending on the listener.

All listeners perform one trial for each of the six types of sounds. Firstly, the listeners performed the experiments for each of the three types of sounds in the anechoic environment and afterwards in the reverberation environment. Provided that the program delivers only one material sound type per trial, it is assumed that this method does not affect participants' response. The program delivers the sounds, one by one, in a randomized order for each stimuli and trial. After hearing the sound through the headphones, the listeners are asked to click on the cell from which they consider the sound has come. Before giving the response, the subjects are allowed to repeat the stimulus. On each trial, the responses are given and recorded by clicking the mouse button pointing the cursor in one position of the cell interface.

RESULTS

The psychoacoustic experiment validates the localization of the source of material sounds via headphones. By means of the Head Related Transfer Functions, the sound obtains the effect of 'free field'; this means that the sound appears as if it comes from the surrounding instead of from the headphones. The acoustical bases of the localization of material sounds processed both in the anechoic environment and in the environment with reverberation are investigated.

Sound source position perception is defined by azimuth angle and distance. The azimuth errors are defined as the angular difference between the defined sound source and the perceived sound source. Distance errors are defined in an analogous way.

The first step when analysing the data is to determine the mean response for each listener in each run. Afterwards, all the mean responses are collected and the global mean for all listeners calculated. Figure 4 illustrates the mean and standard deviation corresponding to material sound localization both for anechoic and for reverberate sounds. The standard deviation ($N-1$) is calculated from the mean responses for each sound in both conditions.

$$sd = \sqrt{\frac{\sum_{k=0}^{N-1}[x(k)-\bar{x}]^2}{N-1}}, \tag{3}$$

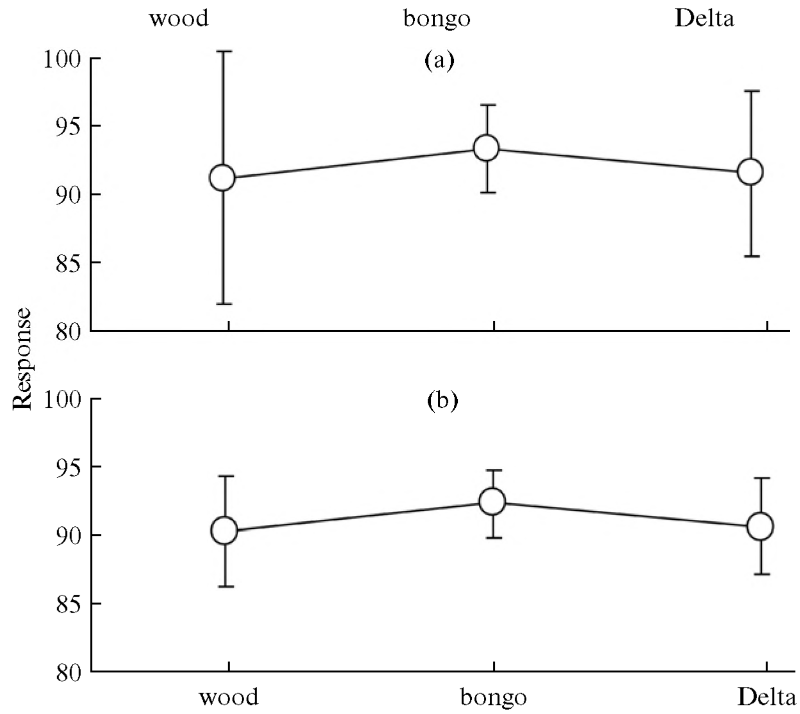

Fig. 4. Average localization responses for ten listeners plotted as a function of material wood, bongo and Delta (click) in %, where (a) represents the results for the anechoic sounds and (b) represents the results of the sounds with reverberation. Errors bars are standard deviation.

where x is the localization, mean $\bar{x}$ represents its average and N represents the total runs number.

Figure 4 shows that there are no great differences between anechoic and with reverberation localization accuracy but great differences appear in the standard deviation. Note that the standard deviations are smaller for the responses corresponding to sounds with reverberation.

Table 2 and Fig. 4 prove that mean values for wood and Delta sounds with reverberation show slight improvements in comparison with the anechoic sounds. This demonstrates that the reverberation effect improves the subjective realism. Again, we can confirm that reverberation effect helps humans on sound localization, at least in distance [18].

On the other hand, it is difficult to reach the same conclusion for the "bongo" sound, in which data differ only in one percent between anechoic and with reverberation sounds.

Also, small differences can be observed in the standard deviations when comparing the anechoic and with reverberation bongo and Delta sounds, but for wood sounds the standard deviation dramatically increases. It is observed that the deviation for the anechoic wood sound is more than twice higher than for the environment with reverberation. This means that the reverberation effect provides a higher precision in the perception judgments. On the other hand, the anechoic sounds are, due to their inherent physical properties, poor, leading to troubles in the localization task, especially for the wood material sound. Figure 5 shows the mean results and the differences for the ten listeners which take part in the experiment for wood sound localization. From Fig. 5 we can observe that listener 2 is one of the not successful listeners, he has

Table 2. Mean and minimum values in percentage for all three rounds for anechoic and reverberant situations

Value, %	Anechoic			Reverberation		
	wood	bongo	Delta	wood	bongo	Delta
Maximum	97.1	96.6	95	96.7	95.4	**94.9**
Mean	89.87	93.38	88.61	91.63	92.83	**91.11**
Minimum	71.5	88.4	80.5	83	87.7	**83.1**
Mean *sd*	**8.8**	**3.1**	**5.8**	**4**	**2.5**	**3.5**

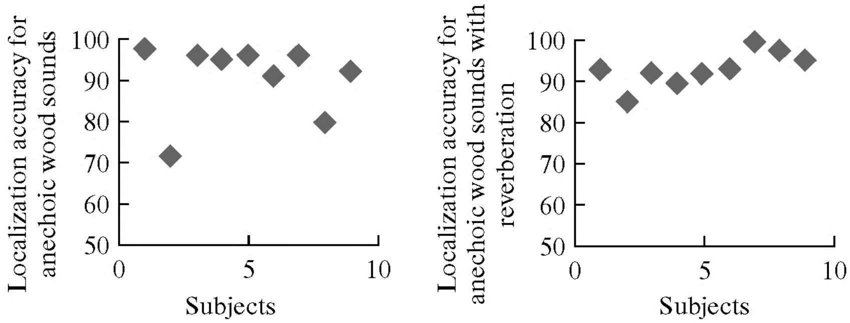

Fig. 5. Localization accuracy for wood sound (anechoic and with reverberation). The x-axis represents the averaged localization data by subjects and y-axis represents the localization accuracy in %.

troubles in the localization task in both the anechoic and with reverberation sounds. For this user, the wood sound was confusing and less clear. Another interesting conclusion is that the localization accuracy in distance—or depth—is worse than in azimuth for almost all listeners. This fact can be justified because the Interaural Level Difference and Binaural Time Difference give less information regarding distance localization [19]. Also, many non-acoustic properties contribute to the sound localization in distance [20].

In most cases, the listener distraction has an important influence on localization accuracy; after several sounds, some listeners become distracted by the environment surrounding the listener or the external noises such as street noises.

One of the listeners, listener X, develops the experiment to analyze the influence of a long training period on the localization task for anechoic material sounds. The listener X is supposed to carry out this test due to his great abilities for sound localization in an anechoic chamber and with headphones as well as to the low Interaural Time Difference threshold obtained by him in previous experiments. Analyzing localization in distance and azimuth, the listener X obtained better results for azimuth material sound localization. Performances on localization and standard deviation ($N-1$) are presented in Table 3.

Taking into consideration the standard deviation ($N-1$) for listener X in Table 3, we can observe that the localization performances in azimuth are more constant than the localization results in distance. This means that the listener perceives better the direction of the sound source than the distance between the sound source and himself. For example for the Delta sound, the standard deviations ($N-1$) for azimuth and distance localization are very small and similar. In this case, listener X is able to accurately perceive the Delta sound for azimuth and distance with a standard deviation of 2.11%. With the bongo sound source, there is a great variation in the results regarding distance localization; the standard deviation is 11.3%. For the bongo sound source, listener X has troubles with distance localization for the first four runs. We must mention that, for the bongo sound, an improvement in distance localization is perceived in the first six runs.

In Fig. 6, we can observe that, during the first four trials, listener X experiences significant oscillations in perception accuracy for distance between material sounds. From the fifth trial, listener X improves his localization accuracy both in distance and in azimuth.

Listeners are able to represent the sound source position with great accuracy.

The present research has direct implications for the design of environments for Virtual Reality in which the interaction with the objects is presented in a similar way to the human visual interaction in the real world. Also, it may be very useful for the analysis and the development of the acoustical navigation systems for blind people. Firstly, the study shows that sound localization in distance and azimuth does not depend on the type of sound material. It is known that human brain is able to perceive and make decisions about sound localization when sounds are presented through the headphones. The second interesting remark is that nowadays all psychoacoustic experiments are performed in special rooms (anechoic and with reverberation) [8], where the level of reverberation and noises are controlled by the experimenter, but there are no works developed in normal environments, where unexpected noises and distractions are present. In the first case all distractions are controlled; only the tired-

Table 3. Mean localization performances in percentage and standard deviation for wood, bongo and Delta material sound in anechoic environment for listener X

	Azimuth	Distance
Wood, %	***92.18***	***88.85***
sd	3.13	5.91
Bongo, %	***93.01***	***86.71***
sd	3.98	11.53
Delta, %	***93.13***	***93.75***
sd	3.86	3.95

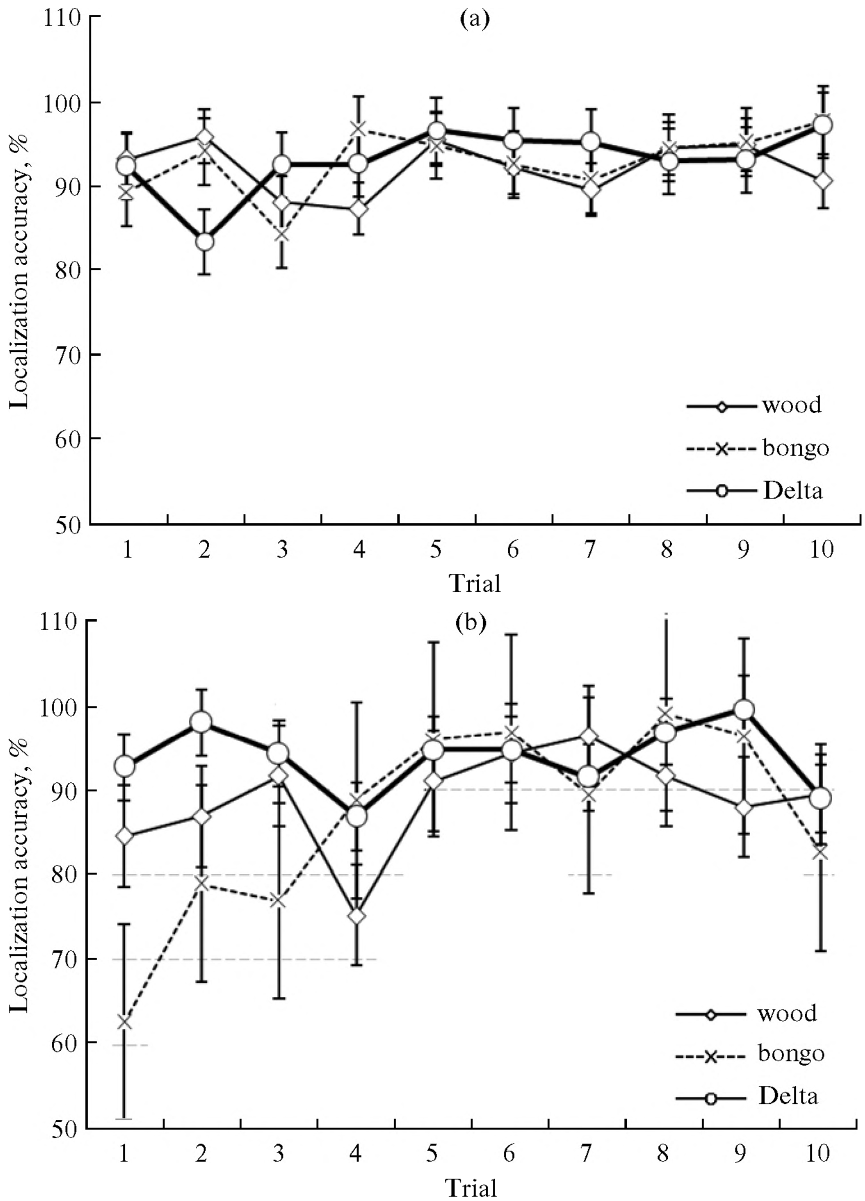

Fig. 6. Listener X perception accuracy for anechoic sounds localization (a) in azimuth and (b) in distance. The x-axes are the trials and the y-axes represent the localization accuracy in %.

ness and the boredom are uncontrolled and strongly affect the results.

CONCLUSIONS

In the present paper, different experiments on localization of sounds coming from different materials are developed and analyzed. In order to create the spatial sensation of the sounds, as if they were coming from the real world, each sound was convolved with Head-Related Transfer Functions. These were measured for each spatial position in a frontal plane by using KEMAR dummy head. Experimental results prove that sound localization in distance and azimuth does not depend on the material sound. Also, the reverberation effect does not significantly increase the localization accuracy. Best results were obtained by the listener X, whose localization results are more constant for direction and distance.

ACKNOWLEDGMENTS

This research was supported by Research Center in Graphic Technology from the Universidad Politècnica de València.

REFERENCES

1. D. S. Brungart and W. M. Rabinowitz, J. Acoust. Soc. Am. **106**, 1465 (1999).
2. D. S. Brungart, I. Nathaniel, and W. R. Rabinowitz, J. Acoust. Soc. Am. **106**, 1956 (1999).
3. H. Bruce and D. Hirsh, J. Acoust. Soc. Am. **31**, 486 (1959).
4. D. I. Shore, S. E. Hall, and R. M. Klein, J. Acoust. Soc. Am. **103**, 3730 (1998).
5. J. C. Kidd and J. H. Hogloben, J. Acoust. Soc. Am., **116**, 1116 (2004).
6. L. Dunai, G. P. Fajarnes, B. D. Garcia, N. O. Araque, and F. B. Simon, Acoust. Phys. **55**, 448 (2009).
7. L. Dunai, G. P. Fajarnes, B. D. Garcia, and V. S. Praderas, Acoust. Phys. **56**, 348 (2010).
8. M. Gröhn, *Proc. Int. Conf. on Auditory Display, Kyoto, 2002.*
9. E. S. Malinina and I. G. Andreeva, Acoust. Phys. **56**, 576 (2010).
10. E. D. Shabalina, N. V. Shirgina, and A. V. Shanin, Acoust. Phys. **56**, 525 (2010).
11. A. Pompey, M. A. Sumbatyan, and N. F. Todorov, Acoust. Phys. **55**, 760 (2009).
12. R. L. Klatzky, D. K. Pai, and E. P. Krotkov, Presence: Teleoperators and Virtual Environments **9**, 399 (2000).
13. M. Aramaki, M. Besson, R. Kronland-Martinet, and S. Ystad, *Proc. 5th Int. Symp. on Comp. Music Model. Retriev. (CMMR 2008), Copenhagen, 2008*, pp. 1–8.
14. W. Gaver, *PhD Dissertation*, Univ. California, San Diego, 1988.
15. N. I. Durlach, A. Rigapolus, X. D. Pang, W. S. Woods, A. Kulkarni, H. S. Colburn, and E. M. Wenzel, Presence: Teleoperators and Virtual Environments **1**, 251 (1992).
16. S. A. Gelfand, *Essentials of Audiology, 3rd ed.* (Thieme Medical Publishers, New York, 2009).
17. J. Jerger, ASHA **4**, 139 (1962).
18. H. Mershon, W. L. Ballenger, A. D. Little, P. L. McMurtry, and J. L. Buchanan, Perception **18**, 403 (1989).
19. D. O. Kim, A. Moiseff, T. J. Bradley, and J. Gull, Acta Otolaryngologica **128**, 328 (2008).
20. P. Zahorik, *Proc. Int. Conf. on Auditory Display, Kyoto, 2002.*