ISSN 1063-7710, Acoustical Physics, 2010, Vol. 56, No. 3, pp. 348–353. © Pleiades Publishing, Ltd., 2010.

ACOUSTIC SIGNAL PROCESSING AND COMPUTER SIMULATION

The Influence of the Inter-Click Interval on Moving Sound Source Localization for Navigation Systems[1]

Larisa Dunai, Guillermo Peris Fajarnes, Beatriz Defez Garcia, Victor Santiago Praderas, and Ion Dunai

Researh Center in Graphic Technology, Universidad Politecnica de Valencia, Camino de Vera s/n, 8L, 46022, Valencia, Spain

e-mail ladu@upv.es

Received September 1, 2009

Abstract—In this paper an analysis of moving sound source localization via headphones is presented. Also the influence of the inter-click interval on this localization is studied. The experimental sound is a short delta sound of 5 ms, generated for the horizontal frontal plane, for distances from 0.5 m to 5 m and azimuth of 32° to both left and right sides with respect of the middle line of the listener head convolutioned with individual HRTFs. The results indicate that the best accurate localization was achieved for the ICI of 150 ms. Comparing the localization accuracy in distance and azimuth is deduced that the best results have been achieved for azimuth. The results show that the listeners are able to extract accurately the distance and direction of the moving sound for larger inter-click intervals.

DOI: 10.1134/S1063771010030139

1. INTRODUCTION

Humans have remarkable ability to perceive their surrounding through hearing. They are able to detect, identify and localize the sound source around them, to roughly estimate the direction and distance of the sound source, the static or moving sounds and the presence of an obstacle or a wall.

Sound source localization has been studied during many years [1]. Lord Rayleigh in his “duplex theory” presented the foundations of the modern research on sound localization [2], introducing the basic mechanisms of localization. Blauert defined the localization as “the law or rule by which the location of an auditory event (e.g., its direction and distance) is related to a specific attribute or attributes of a sound event” [3]. Acoustical cues (interaural cues such as Interaural Time Difference ITD and Interaural Level Difference ILD), torso and pinnae contribute to the human sound localization [4, 5]. Beside that, the Head Related Transfer Functions (HRTFs) are crucial for sound source localization [6]. According to Blauert [3], HRTFs represent the transfer characteristics of the sound source in a free field to the listener’s external ear.

Beside the localization of a static sound source, the moving sound localization plays an important role in the human life [7]. In the case of a moving source, changes in the sound properties appear due to the influence of the sound source speed or to speed of the used program for sound emission.

Several experiments have been carried out on static sound localization using headphones [8], [3] but few for moving sound source localization. In the case of localization via headphones, the sounds are localized inside the head [9]. This is known as “lateralization.” Previous studies [10] showed that sound externalization via headphones can be achieved using individual HRTFs, which help the listener to localize the sound out in space [11, 12]. Due to those HRTFs, the convolutioned sounds are localized as real sounds [13, 14].

Several experiments evaluating the sound localization have been carried out recently. In the first of these experiments [15] the localization of the position of a single sound and a train of sounds was carried out for different inter-click intervals. The initial sound was a monaural delta sound of 5 ms processed by HRTFs filter. The ICIs were varying from 10 ms to 100 ms. The listeners were asked to inform what they listened, the number and the provenience of the listened sound and also if there was any difference between them, evaluating the perceived position of the sound (“Left,” “Right” or “Centre”). It was proven that the accurateness in the response improves with the increase of the length of ICI. Moreover, the train of clicks was localized better than the single click due to the longer time to listen and perceive the sound provenience.

In the second study [16], the real object localization based on sensory system and acoustical signals was carried out via a cognitive aid system for blind people (CASBliP). In this research, the blind users were walking along a 14 m labyrinth based on four pairs of soft columns should localize the columns and avoid them. The average time of sound externalization

[1] The article is published in the original.

and object detection was 3.59 min. The device showed no definitive results due to the acoustical signal speed, which required improvements.

The goal of this study is to analyze how localization of a moving source is influenced by the inter-click interval, and how the listeners localize the moving sounds through headphones.

2. EXPERIMENTS

2.1. Subjects

Nine young subjects with ages between 25 and 30 years and different gender, all of them had normal vision, were involved in the experiments. All participants had normal distance estimation and good hearing abilities. The participants had distance perception compressed relative to the actual distance. They demonstrate a correct perception of the sounds via headphones. The subjects were identified by a number P1-P9.

All subjects participated in previous auditory experiments in the laboratory. Each participant received a description of what was expected of him/her and about all procedure. All participants passed the localization training and tests described below.

2.2. Signal Processing

The distance range for the acoustical module covers from 0.5 m to 5 m, an azimuth of 64°, and 64 sounding pixels per image at 2 frames per second. A delta sound (click) of 2048 samples and sampling rate of 44.100 Hz has been used. To obtain the spatial sounds a delta sound were convolved with Head-Related Transfer Function (HRTF) filter measured for each 1° in azimuth (for 32° left and 32° right side of the user) at each 1 cm in distance.

Recording of directional transfer functions were carried out in an anechoic chamber. The HRTFs measurements system is based on a robotic and acquisition system. The robotic system consist of an automated robotic arm, which includes a loudspeaker, and a rotating chair on an anechoic chamber. A manikin was seated in the chair with a pair of miniature microphones in the ears. In order to measure the transfer function from loudspeaker-microphone as well as for headphone-microphone, the impulse response using Maximum Length Binary Sequence (MLBS) was used. The impulse response was obtained by taking the measured system output circular cross-correlation with the MLBS sequence. The impulse response is given by:

$$h(n) = \Omega_{sy}(n) = s(n)\Phi y(n) = \frac{1}{L+1}\sum_{k=0}^{L-1} s(k)y(n+k), \quad (1)$$

where $y(n)$ is the system output, $s(n)$ is the MLBS and Φ represents the circular cross-correlation.

Because the direct implementation of the Eq. (1) requires a long processing time, we used the equivalent operation of the cross-correlation, convolution passing to the frequency domain. In that case the convolution is a vectorial multiplication:

$$a(n)\Phi b(n) = \frac{1}{L+1}a(-n)b(n). \quad (2)$$

In order to reduce the computational time the Fast Hadamard Transform (FHT) has been used. In that case the impulse response is given by:

$$h(n) = \frac{1}{(L+1)s[0]}P_2\langle S_2\{H_{L+1}[S_1(P_1 y(n))]\}\rangle, \quad (3)$$

where P are the permutation matrices, S—the redimention matrices, and H_{L+1} is the Hadamard matrices of $L+1$ degree.

The output signals (the HRTF) are sampled 22050 Hz and a length of 46 ms (8192 bit). The HRTFs were measured for the horizontal frontal plane at the ear level from 0.5 to 5 m in distance and in azimuth between 32° left and 32° right with respect to the center of the listener head (measurements at every 1°).

2.3. Equipment

For the sound generation and processing, a Huron system with 80 analogue outputs, eight analogue inputs and eight DSPs 56002, and a computer for off-line sound processing has been used.

For the experimental test, SENNHEISER headphones model HD 201 were used to deliver the acoustical information. The model has been selected because it has a good stereo sound and it attenuates the ambient noises; minimum interference with external sounds is desirable in order to obtain best acoustical results.

Matiab 7.0 has been used as experimental software. The resultant graphical sound trajectory for each experiment has been displayed on a separate window and saved for off-line processing. All experiments were done on an ACER Aspire 5610 computer.

3. PROCEDURE

The goal of the experiments is to analyze the localization of a moving sound source via headphones and to see how the inter-click interval (ICI) influences the

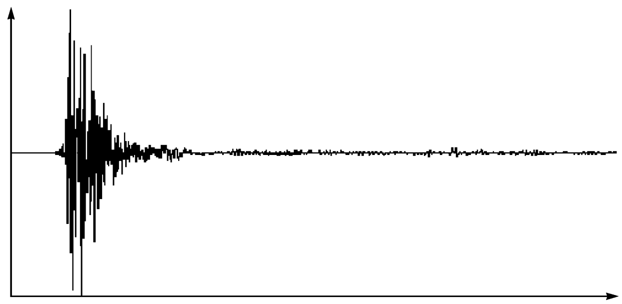

Fig. 1. HRTF wave form of 22050 Hz sample rate, and the length of 46 ms of 8192 bit. On the axe is represented the sound sample.

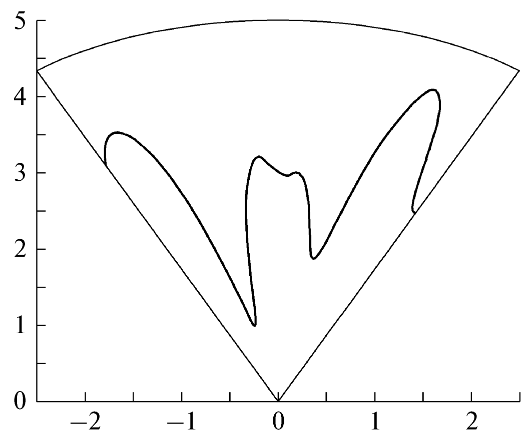

Fig. 2. Sound trajectory, direction from left to right. The *x* axis represents the azimuth where the 0 is the centre of the head, which is 0°. The −2.5 is the −32° at left side of the head and 2.5, respectively, is 32° at the right side of the head. The *y* axis represents the distance from 0 to 5 m.

Fig. 3. Experimental scenario. The user is seated on a chair in front of a computer. Hearing the sound through headphones he should draw the perceived sound trajectory on the paper.

sound localization quality. The comparison between the localization performances enables to evaluate the importance of the inter-click interval parameter for its use in sound localization and acoustical navigation systems.

The movement of the sound source was achieved by switching the convolutioned sound for a frontal plane at the eyes level at increasing distances from 0.5 to 5 m (1 cm increase) and for azimuth between 32° right and 32° left (1° increase) with respect the middle of the head. When the convolved sounds were prepared to be delivered, inter-click intervals of 200 ms, 150 ms, 100 ms, 75 ms, and 50 ms were introduced.

The experiment was carried out during one single session. The session consisted of five runs, one run for each aforementioned ICI.

For the integration of the training exercises in the experiment, four different trajectories were created. The delivered trajectory was selected randomly by the computer when the experiment starts. Figure 2 shows one of the trajectories the sound was running.

The participants were asked to seat comfortably in the chair in front of a computer (see Fig. 3). Before starting the experiment, the training exercises were carried out. The objective and the procedure of the experiment were explained to each individual participant. One sound was delivered for all five ICIs, where the participants were able to see graphically the listened sound trajectory and how sound the trajectory at different ICIs.

Both in the training exercise and the experiment, a sound at a specific ICI was delivered by the computer via headphones. During the experiment, the participants were free to move. Nevertheless, they were required to move the less possible and to be concentrated on the sound, in order to create a plane of the sound route in the imagination. The test was performed both with open eyes and with closed eyes depending on the participant wishes. In the case of the closed eyes, there was a limitation of effects of the visual inputs. Due to this, the participant achieved a better interpretation of the trajectory image.

The participants were asked to carefully hear the sound and draw the listened trajectory in a paper. They were allowed to repeat the sound if it was necessary. All the participants often asked to repeat the sound at least three times. Each participant was supposed to have five trials, one for each ICI. Only one sound trajectory was used per participant for all five ICIs. For all participants the experiment started with the ICI of 200 ms continuing in a decreasing order up to 50 ms.

After the experiment the participants commented the perceived sound trajectory and they compared the heard sound for each ICI.

4. RESULTS

The moving sound source localization is an important factor for the navigation task improvement. The main variables analyzed in this paper were the moving sound source localization and the inter-click interval ICI [200, 150, 100, 75, and 50 ms]. The study analyzes the interaction between these variables in measurements of distance and azimuth.

Generally, no significant differences on the results were registered between participants. However, great difference was found in the sound localization between the large and small inter-click intervals.

The minimum and maximum data of the distance and azimuth are presented in table where, the maximum error in distance is 1.26 m for an ICI of 50 ms and the minimum error was 0.42 m for an ICI of 150 ms, the maximum error in azimuth was 11.4° for an ICI of 50 ms and the minimum 0.71° for an ICI of 150 ms.

Average error results of sound localization in azimuth and distance as a function of the inter-click interval are shown in Fig. 4. Best results have been achieved for large ICI, due to the time needed by the brain to perceive and process the received information. Because the time between two sounds is longer, the sound is perceived as jumping from one position to another from left to right in equal steps. For the ICI of 200 ms, the sound was perceived not so much as a moving sound, but rather as a jumping sound from location to location. However, for the ICIs smaller than 100 ms the sound was perceived as a moving sound from the left to right, but there was enough difference between the original sound trajectory and the perceived one. The participants had great difficulties to perceive the exact distance and azimuth, because the sound was delivered too fast. Moreover, when the sound trajectory had multiple turning points on a small portion of the space, the participants perceived this portion as one turn-return way. Figure 5 represents a specific case, corresponding to one participant, of the moving sound localization at four ICIs. The red colour represents the listened sound trajectory drawn by the participant. The grey colour represents the real sound trajectory drown by the computer. The x axis represents the azimuth where the 0 value is the centre of the head, the negative values are the values at the left side of the head, whereas the values at the right side of 0 represent the azimuth values at the right side of the human head. The –2.5 represent the 32° at left side of the head and 2.5 the 32° at the right side of the head. The y axis represents the distance from 0 to 5 m.

In some cases the participants perceived the sound trajectory near to a straight line when the inter-click interval was 50 ms. Even repeating several times the experiment, the participants were confused regarding the localization of the moving sound. They commented "the sound moves too fast and I feel that is running from left to right in a straight line." Despite listeners were not able to localize the moving sound source at small inter-click intervals as well as they were able to localize the moving sound for large inter-click intervals, they were able to judge about the sound position in azimuth and distance.

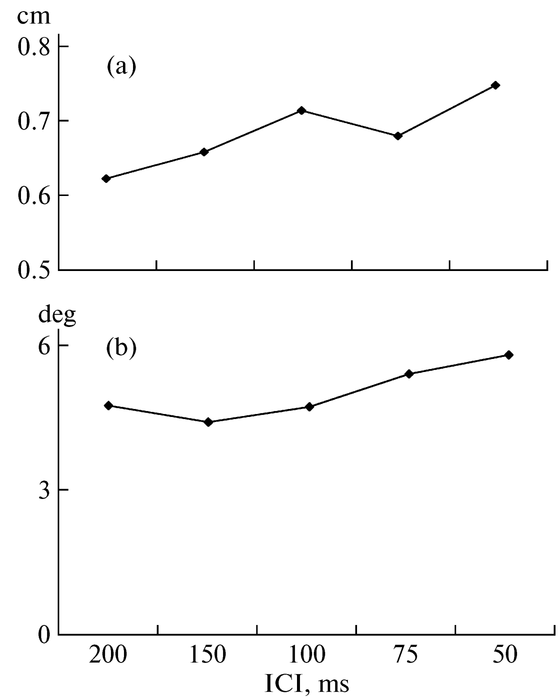

Fig. 4. Average error in distance and azimuth for all participants. In the graphic (a) are represented the average error in distance for all participants for each five inter-click intervals. The small value of 0.63 cm is obtained for the ICI of 200 ms. The average error of the experiment is 0.69 cm. The graphic (b) represent the average error in azimuth, where the small average error of 4.37° was obtained for an inter-click interval of 150 ms. The total average error in azimuth is 4.97°.

Various factors as drawing abilities (how the participants can accurately draw), sound interpretation (how the participants can interpret the heard sounds, by colours, by image etc.), the used hearing methods (with closed or opened eyes), the external noises, etc., influenced the experiment results. Despite all participants were informed about the use of one sound per participant for all ICIs, they draw the trajectories at

Representation of the minimum and maximum error for each five inter-click interval obtained from all participants. The minimum of 0.42 m error in distance is obtained for an ICI of 150 ms, and the maximum of 1.26 m was for 50 ms. In azimuth the minimum value of 0.71° for an ICI of 150 ms and maximum of 9.14° for an ICI of 100 ms

ICI, ms	distance, m min	max	Azimuth, °min	max
200	0.44	0.88	2.1	7
150	0.42	1	0.71	6
100	0.47	1.01	1	9.14
75	0.56	0.85	2.14	14
50	0.5	1.26	1.43	11.4

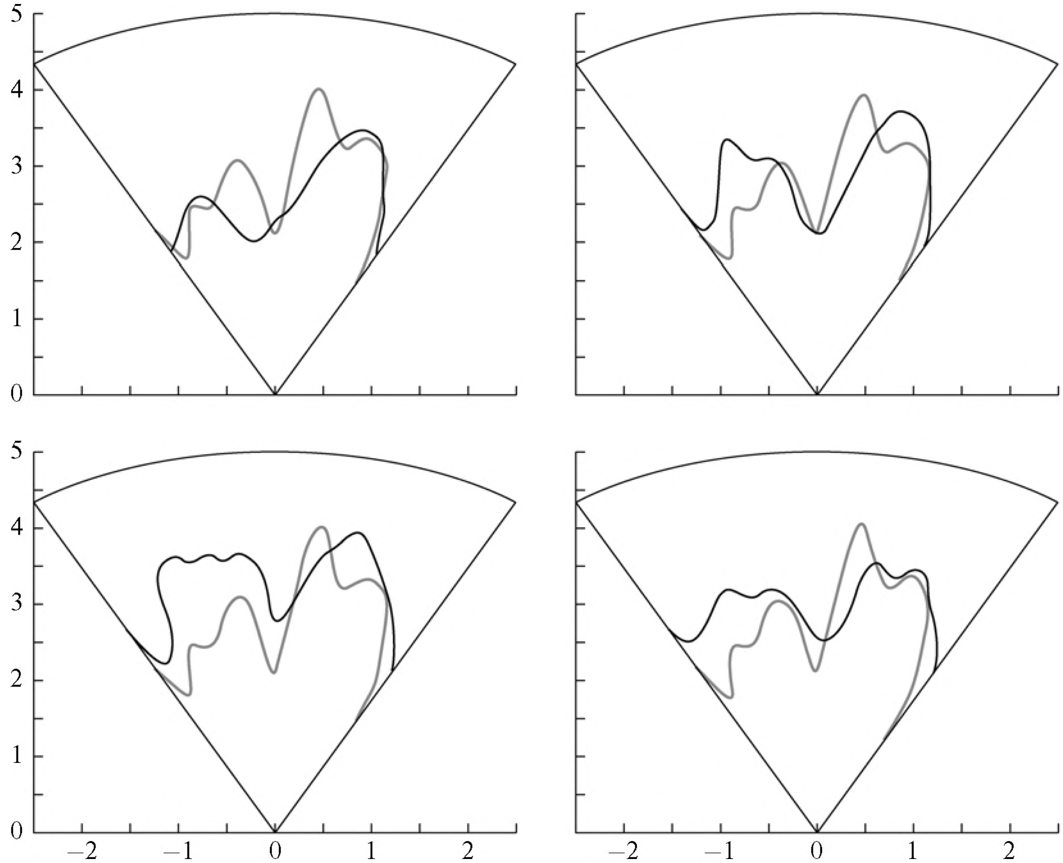

Fig. 5. Sound trajectory for one participant for the ICI of 50 ms, 100 ms, 150 ms, and 200 ms. The red colour represent the heard sound trajectory drown by the participant, the grey colour represent the real sound trajectory drown by the computer. The *x* axes represent the azimuth where the 0 value is the centre of the head, the negative value are the values at the left side of the head and the values at the right side of 0 represent the azimuth values at the right side of the human head. The –2.5 represent the 32° at left side of the head and 2.5 respectively the 32° at the right side of the head. The *y* axis represents the distance from 0 to 5 m.

different distances (see Fig. 5). This error appears because of the participant drawing ability, there is not so easy to interpret graphically what is listened or the image the brain creates if there is not practice on that. For some of participants, great concentration and relaxation was required, to be able to perceive correctly the sounds.

Multiple observations on training sound trajectory were given to participants about how to perceive the sound and to be confident of their answer. Two participants were excluded from the main analysis due to the difficulties in localizing the sound. The participants experienced the moving sound localization as a straight line for all inter-click intervals.

5. CONCLUSIONS

The results showed that when the inter-click interval is larger, the moving sound is better localized such as in distance as in azimuth. The results indicate that the best accurate localization was achieved for the ICI of 150 ms. The analysis indicates that for small ICIs all participants needed more times to repeat the same experiment.

Because the people are quite poor on estimating the distances greater than meter, is the reason why some subjects only heard the stimuli as a “straight line.”

The study has shown that the localization of a moving sound source plays an important role in the human life when is it necessary to direct oneself in the environment.

ACKNOWLEDGMENTS

This study was supported by the European Commission through Cognitive aid System for Blind People (CASBliP) project.

REFERENCES

1. D. S. Brungart, I. Nathaniel, and W. R. Raibiowitz, J. Acoust. Soc. Am. **106**, 1956 (1999).
2. J. W. Strutt, Philos. Mag. **13**, 214 (1907).

3. J. Blauert, Spatial Hearing: The Psychophysics of Human Sound Localization (MIT, Cambridge, MA, 1997).
4. D. S. Brungart and W. M. Rabinowitz, J. Acoust. Soc. Am. **106**, 1465 (1999).
5. H. Bruce and D. Hirsh, J. Acoust. Soc. Am. **31**, 486 (1959).
6. H. Y. Kim, Y. Suzuki, Sh. Takane, and T. Sone, Appl. Acoust. **62**, 245 (2001).
7. Ya. A. Al'tman, V. S. Gurfinkel, O. V. Varyagina, and Yu. S. Levik, Neurosci. Behav. Physiol. **35**, 103 (2005).
8. E. Wenzel, M. Arruda, D. Kistler, and S. Foster, J. Acoust. Soc. Am. **94**, 111 (1993).
9. D. Junius, H. Riedel, and B. Kollmeier, Hearing Researc. **225**, 91 (2007).
10. W. M. Hartmann and A. Wittenberg, J. Acoust. Soc. Am. **99**, 3678 (1996).
11. A. Kulkani and S. H. Colburn, Nature **396**, 747 (1998).
12. G. Versenyi, Acoust. Sci. Technol. **28**, 244 (2007).
13. D. J. Kistler and F. L. Wightman, J. Acoust. Soc. Am. **91**, 1637 (1996).
14. E. M. Wenzel, Presence Telop Virt. Environ. **1**, 80 (1992).
15. L. Dunai, F. G. Peris, B. G. Defez, A. N. Ortigosa, and S. F. Brusola, Appl. Phys. J., No. 3, 448 (2009).
16. L. Dunai, F. G. Peris, B. G. Defez, and A. N. Ortigosa, Living All Eur. Conf. (2009).