Using 3d sound to track one of two non-vocal alarms

Marie Rivenez[1], Guillaume Andéol[1], Lionel Pellieux[1], Christelle Delor[1], Anne Guillaume[1]

[1]Département Sciences Cognitives
IMASSA BP 73
91220 Brétigny sur Orge Cedex
France

`mrivenez@imassa.fr`

ABSTRACT

The sound environment of present day cockpits is extremely dense and pilots experience a constant auditory overload. Todays technology enables us to virtually spatialize sound data using 3D sound. One way of limiting the auditory overload is to spatialize sound data in the headsets. Recent studies in fact show an improvement in the capacity to extract a simultaneous target verbal message from masker messages with separation of the messages in azimuth or in elevation. To our knowledge, the benefit of spatial separation of non-verbal data such as alarms has not been studied. The aim of this study was to evaluate the contribution of spatial separation in azimuth and in elevation to the ability to extract and track a non-vocal alarm type sound sequence simultaneously with a sequence of the same type. We used a detection paradigm with temporal irregularity in a target sequence interleaved with a distracting sequence. We tested the effect of virtual spatial separation between the target and distracting sequences in azimuth (separation of 0°, 10°, 20°, 30°, 40°, 50° and 60°) and in elevation (separation of 0°, 10°, 20°, 30°, 40°, 50°, 60° and 70°) using 3D sound. Two sides were explored: the front and left sides for the azimuth parameter; the front and top sides for the elevation parameter. Participants also had to perform a localization task for each sound used in the experiment. The results showed improvements of temporal irregularity detection performance as the separation in azimuth or in elevation increased. For the azimuth parameter, this improvement was enhanced in the front side as compared with the left one. No effect of the side was observed for the elevation parameter. Performance improvement with spatial separation seemed to relate to the target and mask sequence spatial separation for the azimuth parameter but not for the elevation one as expressed by the localization performances.

1. INTRODUCTION

Fighter pilots in their aircraft experience many perceptive and cognitive constraints linked to the mobilization of attentional resources, strong time pressures and a particularly noisy sound environment (cabin noise, radio messages, verbal interactions, vocal and non-vocal alarms). The large amount of information which the pilot has to process simultaneously (visual information, visual and auditory alarms, verbal communications…) mean that he must mobilize all his cognitive resources in a short space of time. One of the major problems faced by pilots is that of the simultaneity of alarms. For example, when a fault occurs, it may trigger off a series of alarms, each one linked to a different aspect of the fault. Also, fairly urgent alarms with no connection to each other can sound simultaneously. It then becomes difficult for the pilot to extract each alarm in order to identify it. One solution for facilitating the extraction of sound data in a complex auditory environment is to virtually spatialize the different sources. The virtual spatialization of auditory data (3D sound) is made possible by the new technologies which can be integrated in to pilot headsets. 3D sound has a number of advantages: it improves detection performance, increases situational awareness and reduces workload [1] [2].

Compared with frequency, spatial cues have for long time been considered as of little interest in the extraction of a sound source from an auditory scene [3]. More recently, a considerable number of experimental data gathered in the aerospace domain have tended to give them a more important role [1]. These studies suggest that the capability of extracting an information from surrounding noise (noise bands, simultaneous verbal messages…) is improved by the virtual spatial separation of co-occurring information via 3D sound. In an operational context (flight simulator), Begault & Pittman (1996) [4] found an acceleration of 500ms of virtually spatialized auditory target detection times compared to non spatialized targets. This advantage of 3D sound seems to result from the spatial separation in azimuth and in elevation of the sources.

1.1. Effect of separation in azimuth

A separation in azimuth greatly improves the intelligibility of segments of speech or phrases. By measuring the proportion of words correctly recalled from a target message run simultaneously with a distracting message, Drullman & Bronkhorst (2000) found an improvement of 43% by introducing a separation of 90% between the messages on the azimuth plane using HRTF's (head related transfer functions) either individualized or generic. This phenomenon was confirmed in a flight simulator by several different studies [6] [7] [2]. Virtual spatial separation also has an effect on speech reception threshold: a difference between a target message and a masker message (filtered in frequency bands in order to minimize energetic masking) of 90° in azimuth improves the intelligibility of the target message by 18dB [8].

Several cues have been proposed in order to explain the effect of spatial separation in azimuth on the ability to extract an information from surrounding noise: monaural intensity cues, binaural cues (Inter Level Difference, ILD, and Inter Time Difference, ITD) and, more recently, perceived location cues. Only the effect of perceived location cues makes a true link between the improvement in performance and the spatial separation as perceived by the listener.

A certain amount of research has shown that the effect of a difference in azimuth between information (phrases, double vowels) on the ability to identify them seems mainly due to the difference in ILD between this information (see for example [9] [10] [11]). In reality, it seems that the critical index is not the ILD but the difference of intensity between the target information and the distracting information (Target-to-Noise Ratio: TNR) in the "best ear". Several authors have in fact observed the same intelligibility performances by presenting target and masker messages binaurally with a difference of ILD and by presenting the "best ear" monaurally [11] [12] [13] [14] [15]. This effect was replicated with non verbal stimuli [16].

The use of interaural physical differences also seems to be involved in the improvement of the performances of extraction of a sound information via a binaural Equalization-Cancellation mechanism (EC theory [17] [18] [19]. Initially, it seems that this binaural process equalizes (equalization) the two inputs of the distracting information (right ear and left ear) in order to reduce the interaural differences (ILD & ITD). It then subtracts the two equalized inputs (cancellation) thereby improving the TNR. This mechanism would operate in each frequency band [17] but could only be used if the spectro-temporal contents of the target and distracting information were superimposed [12], which is not always the case for non-vocal alarms which are generally composed of sequences of intermittent sounds.

Finally, it seems that perceived location occurs rather in tasks involving attention focusing on an information [20] and/or where there is only a small spectral overlap between target and masking information [12]. In this respect, the recent study by Shinn-Cunningham et al. (2005) [12] in particular observed a limit in the extraction power of the "best ear" listening process. The authors show that, for an improvement of 12.5 dB with a location difference of 90° between a target message and a masker message, the TNR in the "best ear" accounted for 7.5dB of improvement, the remaining 5dB being explained by the difference of perceived location. The assumption of the use of differences in perceived location is supported by the data provided by Gallun et al. (2005) [21]. The authors observed an improvement in intelligibility performances with the difference of perceived location between target and masker messages for a constant TNR in the "best ear".

1.2. Effect of separation in elevation

To our knowledge, only Worley & Darwin (2002) [22] have looked at the effect of separation in elevation on the ability to track an auditory information. The authors showed the possible use of elevation for tracking a message. They simultaneously presented the target phrase "Could you please write the word *speech* down now" and the distracting phrase "You will also hear the sound phrase this time". The participants had to report the word *speech* contained in the phrase "Could you please write". The experiment was carried out in free field conditions (the messages were presented via loudspeakers). With a difference of elevation of 31° between phrases, in 95% of cases, the listeners attributed the target word to the phrase having the same elevation (performances being random for a difference of 2.5°).

1.3. General aim

The general aim of this study was to measure the effect of spatial separation in azimuth and in elevation of sound information on the ability to extract and track a non-vocal alarm type sound sequence.

Previous studies showed the importance of virtual spatial separation in azimuth and in elevation for the intelligibility of a verbal message. However, it seems that no study has looked at the use of this cue for extracting and tracking a non-vocal sound alarm simultaneously with other information of the same type. In order to simulate the extraction and tracking of a non-vocal alarm, we used a temporal irregularity detection task in a regular target sequence interleaved with an irregular mask sequence. In order to isolate the effect of spatial separation from that of the spectrum, the target and mask sequences had the same spectral content (before being spatialized). The temporal irregularity was inserted into the target sequence and consisted of a phase delay applied to one of the sounds making up the target sequence (figure 1). The separation in azimuth or in elevation of the target and mask sequences was manipulated.

In the light of previous studies carried out principally in the field of speech intelligibility, we considered that we should observe an improvement in irregularity detection performance with the separation in azimuth and in elevation of the target and mask sequences. The minimum angle of separation between target and mask information, necessary for tracking, would serve as an index for the design of spatialized sound alarms in cockpits.

2. GENERAL METHOD

Each trial lasted approximately 11 seconds and was composed of two successive, regular target sequences separated by 1200ms. A temporal irregularity was inserted into the first or the second target sequence. Each target sequence was interleaved with a mask sequence (there was no superimposition of the sounds of the two sequences). Figure 1 shows the sequence design principle.

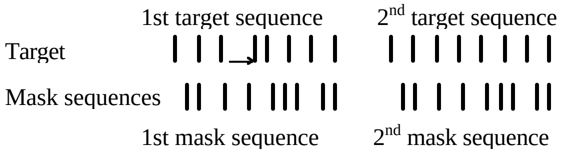

Figure 1. *Schematic view of sequences used (target and mask sequences). Each bar represents a sound of 50ms. In this example the temporal irregularity (represented by an arrow) was inserted into the first target sequence.*

The target sequence being composed of 10 white noise bursts of 50ms (including rise and fall times of 10ms) presented at a rate of 500ms IOI (Inter-onset-Interval). There were 12 different mask sequences, composed of 11 white noise bursts on average, succeeding each other at a rate of 387ms IOI (± 233ms average). During a given test, the mask sequence was the same for the first and second complex sequence. The mask sequence always started 50ms after the target sequence.

The temporal irregularity consisted of a phase shift of 55ms (see figure 1) of a sound from the target sequence (11% of the rate of the target sequence, corresponding to a value much higher than the temporal irregularity detection thresholds see [23] [24] [25]). This shift could be applied to the 5th, 6th, 7th or 8th sound in the first or second target sequence. There were therefore 8 possibilities, each being repeated 3 times, resulting in 24 trials.

A control condition was added in which 24 target sequences were presented singly (without mask sequence) in order to ensure that the participants were suitable to carry out the temporal irregularity task.

The sequences were generated by a TDT RX6 processor controlled by a PC running Windows XP using Matlab 7.0.1. Programming. The sounds were emitted through converters integrated into the TDT processor at 48828 Hz and 24 bits. The responses of the participants were recorded using an external pendant control unit connected to the processor.

The target and mask sequences were presented binaurally at 60dBA via a Beyer TD990 stereophonic headphone (open headphone with diffuse field equalization). The emission level of the target and mask sequences were adjusted by measuring the acoustic pressure (in dBA) supplied by the headphone simulating a sound source of the same spectral content as the sounds making up the sequences, at 140cm from the head and facing (0° azimuth, 0° elevation[1]). This measurement was carried out using artificial ear BK4153.

3. EXPERIMENT 1

The aim of Experiment 1 was to measure the effect of the separation in azimuth of the target and mask sequences on the temporal irregularity detection performances in the target sequence. Two sides were explored: the left and the front side of the listeners.

3.1. Method

3.1.1. *Material*

We used the 24 trials carried out in accordance with the principles described in the general method. The target sequence position was 340° (20° toward the left side of the participant: frontal position) or 250° (110° toward the left side of the participant: left position) (see Figure 2). For the 340° target position, the mask sequence could be presented at 340°, 350°, 0°, 10°, 20°, 30° or 40° in azimuth. For the 250° target position, the mask sequence could be presented at 250°, 260°, 270°, 280°, 290°, 300° or 310° in azimuth. These mask positions led to spatial separations between the target and the mask sequences of: 0°, 10°, 20°, 30°, 40°, 50° or 60°. Elevation was constant and fixed at 0°.

Virtual spatialization was achieved using HRTF's measured in free field on a Neumann KU81i dummy head fitted with mouldings of human pinnas. The signals were emitted by a Forstex 103 broad band loudspeaker located at 140cm from the centre of the dummy head. Impulse response was determined by sending the loudspeaker a signal composed of a pseudo random binary sequence of maximal length (Maximal Length Sequence, MLS). The sequences were of order 13 and an average was taken of 4 stabilised responses. The processor sampling frequency being 48828 Hz, the duration of a sequence was 168ms.

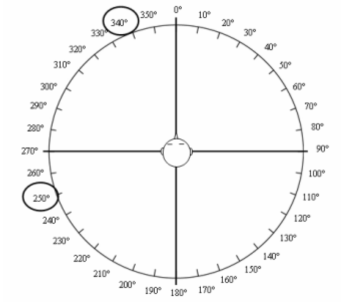

Figure 2. *Representation of the target positions (circles) in azimuth used in Experiment 1. This display was also used to collect participant responses for the localization task.*

---

[1] The 0° in azimuth corresponded to the interaural axis of the dummy head used for the HRTF measurements. The 0° in elevation corresponded to the axis perpendicular to the boom to which the dummy head used for the HRTF measurements was attached.

3.1.2. Procedure

The participant's task consisted of determining whether the temporal irregularity was present in the first or second target sequence.

All the participants satisfied the control condition before the experimental condition. The control condition was composed of 48 trials (24 for each of the two target positions), preceded by 8 practice trials. Each participant then took part in the experiment which consisted of 336 trials presented in a random order. The experiment was carried out in four blocks of approximately 17 minutes, preceded by a practice phase of 8 trials. For half of the participants, the target position was 250° in the first two blocks and 340° in the last ones. For the other half, the target position was 340° in the first two blocks and 250° in the last ones.

The location capabilities of the listeners were measured at the end of the experiment. The participants had to listen to all the bursts used in the experiment for creating the target and mask sequences and locate them. The bursts were repeated 8 times with an IOI of 500 ms. As in the temporal irregularity detection task, the bursts could be presented on two spatial sides (left or front side). Furthermore, they could be presented at 7 locations for each area. For the left side, burst position was 250°, 260°, 270°, 280°, 290°, 300° or 310° in azimuth. For the front side, burst position was 340°, 350°, 0°, 10°, 20°, 30° or 40° in azimuth. Each condition was repeated three times. At each presentation, participants had to locate the burst by indicating its position on a diagram representing a head seen from above, inside a circle on which points were positioned every 10° of azimuth (see Figure 2).

3.1.3. Participants

There were five participants in the experiment. Only participants with a normal audiogram without any notable difference between the two ears were retained for analysis. They had no previous specific practice with 3D sound.

3.2. Results & discussion

3.2.1. Temporal irregularity detection

Mean results are presented in Figure 3. We measured the mean percent correct associated with the temporal irregularity detection. These data were analyzed in 2 x 7 (2 target positions x 7 azimuth differences) repeated analysis of variance measures (ANOVA) in which the two factors were measured within subjects.

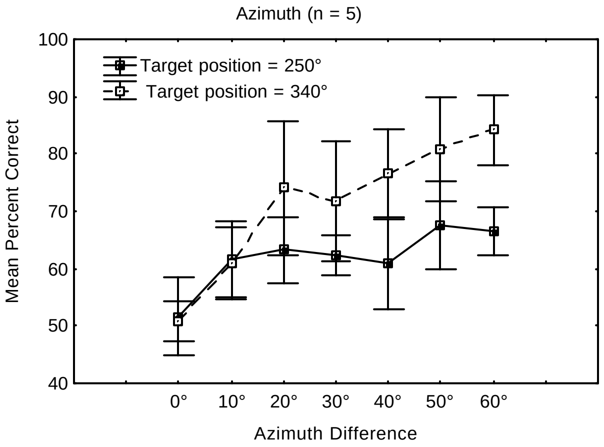

Figure 3. *Mean Percent Correct found in Experiment 1, for the 7 azimuth differences and 2 target positions. Error bars show the standard errors.*

There was a significant effect of the azimuth difference (F(6,24) = 8.43; $p < .0001$) on the mean percent correct: performances improved as the azimuth difference increased (they went from 51% with no separation to 84% with a 60° difference). Percent correct started to be above the chance level (50%) when the difference in azimuth was at least 20° (69%) (t(4) = 2.47; $p < .1$). The effect of target position was marginally significant (F(1,4) = 6.60; $p < .1$): performances were better when the target position was 340° (71%) than when it was 250° (62%). There was a tendency toward an interaction between the azimuth difference and the target position: the effect of azimuth was descriptively larger when the target position was 340° than when it was 250°. For the 250° target position, there was an improvement of 15 points of percentage with a 60° separation while this improvement was 33 points of percentage for the 340° target position. However, this interaction was not significant.

Localization signed errors were analyzed in a 2 x 7 (2 zones x 7 locations) repeated ANOVA measure. Mean signed errors are presented in Figure 4. A negative error means that the azimuth was underestimated while a positive error means an overestimation of the azimuth. The closer to 0 the error was, the better the localization accuracy was. There was a significant interaction between the spatial zone and the localization of the burst on the localization errors (F(6,24) = 35.57; $p < .0001$). Negative errors were larger when the burst was on the front side (-174°) than when it was on the left side (7°) for the first 2 locations (340°, 350° for the front side and 250°, 260° for the left side) (F(1,4) = 25.28; $p < .01$). The reverse was observed for the other burst locations: positive errors were larger when the burst was on the front side (72°) than when it was on the left side (-26°) (F(1,4) = 58.92; $p < .005$).

3.2.2. *Localization task*

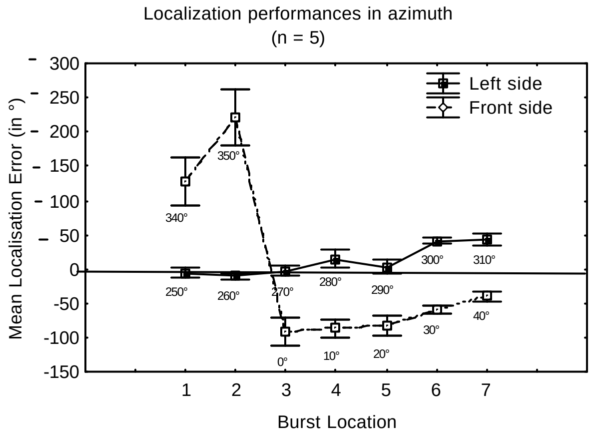

Figure 4. *Mean localization signed error observed in Experiment 1, for the 2 explored sides and the 7 burst localizations in each zone (the exact burst localization is indicated below the error bars). Error bars show the standard errors.*

3.2.3. *Virtual and perceived spatial separation*

In order to investigate the factors involved in the temporal irregularity detection task, we computed correlation analyses between the percent correct and the different parameters involved in the experiment: localization errors, virtual azimuth difference (the one we manipulated using 3D sound) and perceived azimuth difference. Correlations were analyzed with the 14 positions used in the experiment (250°, 260°, 270°, 280°, 290°, 300°, 310°, 340°, 350°, 0°, 10°, 20°, 30° and 40°).

There was no correlation between the performances in the temporal irregularity detection task and the localization errors ($r = 0$). There was a significant and strong correlation between the virtual azimuth difference and the percent correct ($r = 0.73$; $p < .05$): percent correct improved as the spatial separation in azimuth increased as observed in the ANOVA analyzing the percent correct. However, the correlation between the perceived azimuth difference and the percent correct was stronger ($r = 0.82$; $p < .05$) than the one with the virtual azimuth difference.

These data suggest that the determining factor involved in the temporal irregularity detection performances may not be the localization accuracy but the perceived spatial separation between the target and the mask sequence. In order to assess this hypothesis, we carried out a 2 x 6 repeated measure ANOVA on the perceived azimuth difference between the target and the mask sequences. This variable was computed for each participant and each experimental condition (2 target positions x 6 azimuth difference [we excluded the 0° difference because of the absence of variability of this variable]).

Results are presented in Figure 5. The ANOVA revealed a significant effect of the target position on the perceived azimuth difference ($F(1,4) = 14.84$; $p < .05$). The perceived azimuth difference was larger when the target position was 340° (117° on average) than when it was 250° (21° on average). There was no other significant principal effect or interaction.

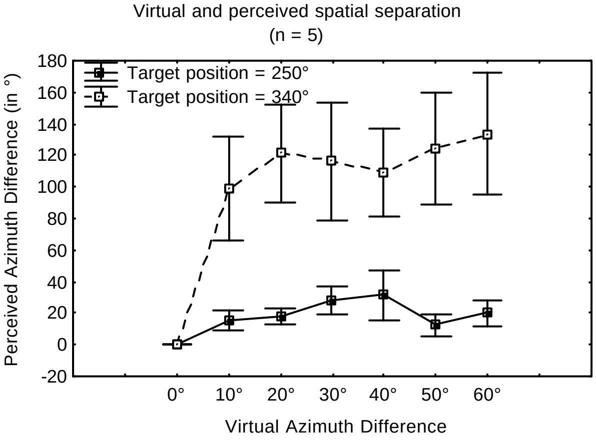

Figure 5. *Mean perceived azimuth difference found in Experiment 1, for the 7 virtual azimuth differences and 2 target positions. Error bars show the standard errors.*

4. EXPERIMENT 2

The aim of experiment 2 was to measure the effect of the separation in elevation of target and mask sequences on performance of detection of a temporal irregularity in the target sequence. We explored two sides: the top and the front side of the listeners.

4.1. Method

4.1.1. *Material*

We used the same material as that for Experiment 1. The only difference was in the spatial configuration of the target and mask sequences. The target sequence was presented at 80° (front side) or 10° (top side) in elevation (Figure 6). For the 80° target position, the mask sequence could be presented at 80°, 90°, 100°, 110°, 120°, 130°, 140° or 150° in elevation. For the 10° target position, the mask sequence could be presented at 10°, 20°, 30°, 40°, 50°, 60°, 70° or 80° in elevation. These mask positions led to spatial separations between the target and the mask sequences of: 0°, 10°, 20°, 30°, 40°, 50°, 60° or 70°. The azimuth was fixed at 10° regardless of the elevation.

4.1.2. Procedure

The procedure was the same as that for Experiment 1 except that there were 384 trials because of the two additional spatial separation conditions.

For the localization task, a head seen in profile was used in order to enable the participants to indicate the position of the bursts (see Figure 6). The head was shown inside a circle on which points were represented every 10° in elevation.

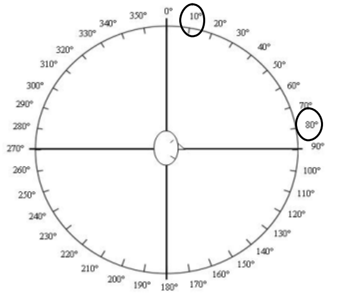

Figure 6. *Representation of the target positions (circles) in elevation used in Experiment 2. This display was also used to collect participant responses for the localization task.*

4.1.3. Participants

Five persons who had not participated in Experiment 1 took part in the experiment. Only participants with a normal audiogram without any notable difference between the two ears were retained for analysis. They had no previous specific practice with 3D sound.

4.2. Results

Mean results are presented in Figure 7. As in Experiment 1, the mean percent correct associated with the temporal irregularity detection was measured. These data were analyzed in a 2 x 8 (2 target positions x 8 elevation differences) repeated ANOVA measures in which the two factors were measured within subjects.

There was a significant effect of the elevation difference (*F*(7,28) = 9.80; $p < .0001$): percent correct improved as the difference in elevation increased (they went from 48% with no separation to 83% with a 80° difference). Performances started to be above the chance level (50%) as the elevation separation was superior or equal to 20° (71%) (*t*(4) = 3.16; $p < .05$). There was no effect of target position: percent correct were not different when the target position was 80° (74%) and when it was 10° (75%). There was no other significant effect or tendency.

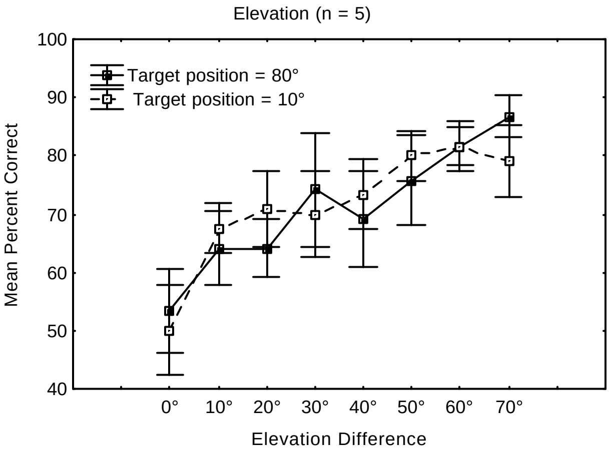

Figure 7. *Mean Percent Correct found in Experiment 2, for the 8 elevation differences and 2 target positions. Error bars show the standard errors.*

4.2.1. Localization task

Only four over the five participants performed the localization task. Localization signed errors were analyzed in a 2 x 8 (2 zones x 8 locations) repeated ANOVA measure . Mean signed errors are presented in Figure 8 (a positive error meant an overestimation of the elevation while a negative error meant that the elevation was underestimated). The average error was 142°. There was an effect of the spatial zone (F(1,3) = 38.23; p < .01): positive errors were larger when the burst position was 10° (172°) than when it was 80° (113°). There was also a significant effect of the localization of the burst (*F*(7,21) = 3.14; $p < .05$). In order to assess the specific significant differences, we performed post hoc analysis as we had no hypothesis about this factor. A Scheffé test revealed no significant difference.

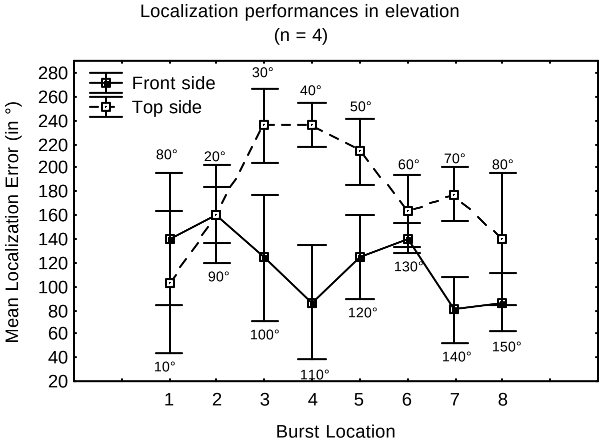

Figure 8. *Mean localization signed error observed in Experiment 2, for the 2 explored sides and the 8 burst localizations in each zone (the exact burst localization is indicated below or above the error bars). Error bars show the standard errors.*

4.2.2. Virtual and perceived spatial separation

As in Experiment 1, we analyzed the correlations between the percent correct, the localization errors, the virtual elevation difference and the perceived elevation difference between the target and the mask sequences. Correlations were analyzed with the 16 positions used in the experiment (90°, 100°, 110°, 120°, 130°, 140°, 150°, 10°, 20°, 30°, 40°, 50°, 60°, 70° or 80°). There was no significant correlation between the performances and the virtual elevation difference ($r = -.04$) or the perceived elevation difference ($r = .26$). As expressed in Figure 9, the perceived elevation difference did not vary with the virtual elevation difference. However, there was a significant and strong correlation between the performances and the virtual elevation difference ($r = .92$; $p < .05$). Performances improved as the spatial separation in elevation increased as observed in the ANOVA analyzing the percent correct.

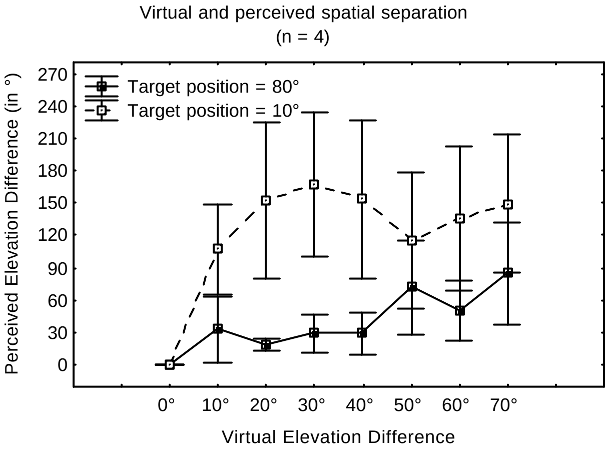

Figure 9. *Mean perceived elevation difference found in Experiment 2, for the 8 virtual elevation differences and 2 target positions. Errors bars show the standard errors.*

5. DISCUSSION

Experiment 1 results support a facilitating effect of the spatial separation when tracking a target sound sequence concurrent with a mask sound sequence. We found that 20° was the minimum azimuth separation to track a sound sequence. Below this limit, participant performances did not differ from the chance level. We also observed that this separation in azimuth advantage seemed to depend on the position of the sequence to be tracked. When the target sequence was on the left side of the listener, the performance improvement with azimuth separation was smaller than when the target sequence was in front. This could be due to spatial sensitivity which is worse on the side than in front of the listener. The minimum audible angle is 3.6° in front while it is 9.2°on the left side [26]. However, the localization performances obtained in Experiment 1 did not indicate such an effect. Participant's errors were larger in front than on the left. This may be due to the large amount of front-back confusion. Furthermore, the absence of correlation between the percent correct in the temporal irregularity detection task and the signed error indicated that localization accuracy may not be a relevant explanation for the side effect. Our data suggest that perceived spatial separation may be a better hypothesis to explain the interaction between the target position and the azimuth difference. When the target was in front, the perceived difference in azimuth between the target and the mask sequences was larger leading to a better discrimination of the two sequences than when the target was on the left.

The results of Experiment 2 showed an improvement of sound sequence tracking when the target and the mask sequences were presented with a different elevation. As for the azimuth dimension, we observed that 20° was the minimum angle necessary to distinguish the target and the mask sequences. There was no effect of the position of the target or interaction between the target side and the elevation difference. Localization errors revealed very poor performances in the participant's ability to localize the sound sequences. The average localization error in the elevation dimension was 142°. Listeners tended to perceive the sounds at the back while the sounds were presented above or in front of them. Furthermore, the absence of correlation between the performances and the localization errors or perceived elevation difference suggest that improvement in tracking the target sequence with the elevation difference may not be due to the spatial separation of the sequences. The effect of the virtual elevation difference between the target and the mask sequence on the ability to track the target sequence may be explained by the spectral differences induced by the HRTF applied to the sounds. The absence of spatial separation effect on tracking the target sequence could be due to the used of non-individualized HRTF. Worley & Darwin (2002) [22] also found that, using non-individualized HRTF to spatialize two simultaneous messages, voice tracking improvement with elevation difference was slightly less marked than with individualized HRTF. The percentage of attribution of the target word to the phrase having the same elevation peaked at 75% with a difference of 40° or 80° between the messages. It seems that their results were due to the poor location capability using non-individualised HRTF's.

6. CONCLUSION

Our data on the effect of azimuth support the idea that spatial separation in azimuth may be a relevant presentation for the design of future non-vocal alarms even using non-individualized HRTF for the spatialization. A minimum angle of 20° in azimuth between two simultaneous non-vocal alarms is necessary in order to allow tracking one of the two alarms. Additional experiments using individualized HRTF for spatializing the alarms are required in order to assess the advantage of the elevation dimension.

Acknowledgements

This work was founded by the Délégation Générale pour l'Armement (CO 07co803 PEA 010801).

7. REFERENCES

[1] N. Iyer, B. D. Simpson, & D. S. Brungart, "An overview of spatial auditory display," Cognitive Processes, vol. 7, no. 1, pp. 148, 2006.

[2] R. L. McKinley, M. A. Erickson, & W. R. D'Angelo, "3-dimensional auditory display: development, applications, and performance," Aviat. Space Environ. Med., vol. 65, no. 5, pp. A31-8, 1994.

[3] A. S. Bregman, Auditory Scene Analysis: The perceptual Organisation of Sounds, MIT Press, Cambridge, MA, 1990.

[4] D. R. Begault, & M. T. Pittman, "Three-dimensional audio versus head-down traffic alert and collision avoidance system display," International Journal of Aviation Psychology, vol. 6, no. 1, pp. 79-93, 1996.

[5] R. Drullman, &, A. Bronkhorst, "Multichannel speech intelligibility and talker recognition using monoral, binaural, and three-dimentional auditory presentation", J. Acoust. Soc. Am., vol. 107, pp. 2224-2235, 2000.

[6] J. A. MacDonald, J. D. Balakrishnan, M. D., Orosz, & W. J. Karplus "Intelligibility of speech in a virtual 3-D environment". Human Factors, vol. 44, no 2, pp. 272-286, 2002.

[7] K. M. McAnally, R. Martin, J. Doman, G. Eberle, & S. Parker, "Detection and identification of simultaneous communications in a simulated flying task", in Proc. of the Human Factors and Medicine Symposium: Improving Audio Effectiveness, Amersfoort, Netherlands, 11-13 April, 2005.

[8] T. L. Arbogast, C. R. Mason, & K. J. Kidd "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am., vol. 112, no. 5, pp. 2086-2097, 2002.

[9] R. W. Hukin, & C. J. Darwin, "Effects of contralateral presentation and interaural time differences in segregating a harmonic from a vowel," J. Acoust. Soc. Am., vol. 98, no. 3, pp. 1380-1387, 1995.

[10] W. R. Drennan, S. Gatehouse, & C. Lever, "Perceptual segregation of competing speech sounds: the role of spatial location," J. Acoust. Soc. Am., vol. 114, pp. 2178-2189, 2003.

[11] B. G. Shinn-Cunningham, J. Schickler, N. Kopco, & R. Litovski, "Spatial unmasking of nearby speech sources in a simulated anechoic environment," J. Acoust. Soc. Am., vol. 110, no 2, pp. 1118-1129, 2001.

[12] V. Best, E. Ozmeral, F. J. Gallun, K. Sen, & V. G. Shinn-Cunnigham, "Spatial unmasking of birdsong n human listeners: Energetic and informational masking factors," J. Acoust. Soc. Am., vol. 118, no. 6, pp. 3766-3773, 2005.

[13] J. F. Culling, M. L. Hawley, & R. Y. Litovsky, "The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources," J. Acoust. Soc. Am., vol. 116, no. 2, pp. 1057-1065, 2004.

[14] G. J. Kidd, C. R. Mason, T. L. Rotha, P. S. Deliwara, "Release from masking due to spatial separation of sources in the identification of non-speech auditory patterns," J. Acoust. Soc. Am., vol. 104, pp. 422-431, 1998.

[15] M. L. Hawley, R. Y Litovsky, & J. F. Culling, "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," J. Acoust. Soc. Am., vol. 115, no. 2, pp. 833-843, 2004.

[16] M. Turgeon, A. S. Bregman, & P. A. Ahad, "Rhythmic masking release: contribution of cues for perceptual organization to the cross-spectral fusion of concurrent narrow-band noises," J. Acoust. Soc. Am., vol. 111, no. 4, pp. 1819-1831, 2002.

[17] J. F. Culling, & Q. Summerfield, "Perceptual separation of concurrent speech sounds: absence of across-frequency grouping by common interaural delay," J. Acoust. Soc. Am., vol. 98, pp. 785-797, 1995.

[18] N. I. Durlach, "Equalization and cancellation model of binaural masking-level differences", J. Acoust. Soc. Am, vol. 35, pp. 1206-1218, 1963.

[19] B. A. Edmonds, & J. F. Culling, "The spatial unmasking of speech: evidence for within-channel processing of interaural time delay," J. Acoust. Soc. Am., vol. 117, no. 5, pp. 3069-3078, 2005.

[20] C. J. Darwin, & R. W. Hukin, "Auditory objects of attention: the role of interaural-time difference," J. Exp. Psychol. Hum. Percept. Perform., vol. 25, pp. 617-629., 1999.

[21] F. J. Gallun, C. R. Mason, & G. J. Kidd, "Binaural release from informational masking in a speech identification task," J. Acoust. Soc. Am., vol. 118, no. 3, pp. 1614-1625, 2005.

[22] J. W. Worley, & C. J. Darwin, "Auditory attention based on differences in median vertical plane position," Proceedings of the 2002 International Conference on Auditory Display, Kyoto, Japan, July 2-5, 2002.

[23] R. Brochard, C. Drake, M. C. Botte, & S. McAdams, "Perceptual organization of complex auditory sequences: Effect of number of simultaneous subsequences and frequency separation," J. Exp. Psychol. Hum. Percept. Perform, vol. 25, pp. 1742-1759, 1999.

[24] C. Drake, & M. C. Botte, "Tempo sensitivity in auditory sequences: Evidence for a multiple-look model," Percept. Psychophys., vol. 54, pp. 277-286, 1993.

[25] I. J. Hirsh, C. B. Monahan, K. W. Grant, & P. G. Singh, "Studies in timing 1: Simple patterns," Percept. Psychophys., vol. 47, pp. 215-226, 1990.

[26] Mills, A. W., "On the minimum audible angle," J. Acoust. Soc. Am, vol. 32, 132-134, 1958.