FEASIBILITY OF MULTIPLE NON-SPEECH SOUNDS PRESENTATION USING HEADPHONES

Gaëtan Lorho[1], Juha Marila[2], and Jarmo Hiipakka[1]

Nokia Research Center
[1]Speech and Audio Systems Laboratory
[2]Visual Communications Laboratory
P.O. Box 407, FIN-00045 Nokia Group, Finland
[gaetan.lorho, juha.marila, jarmo.hiipakka]@nokia.com

ABSTRACT

This paper describes a study of listeners' ability to segregate spatially separated sources of non-speech sounds. Short sounds from musical instruments were played over headphones at different spatial positions using stereo panning or 3-D audio processing with Head-Related Transfer Functions. The number of sound positions was limited to five in this study. One, three or five sound items were played to the listener, multiple sounds being presented with four different onset times from simultaneous to successive replay. The subjects had to spatially discriminate one sound item, i.e. identify a given instrument and find its position. Performance was assessed by measure of response time and error-rate. A preference grading was also included in this test to compare the two headphone presentation techniques employed.

1. INTRODUCTION

Spatial organisation of information is widely used in graphical user interfaces where multiple visual objects can be easily displayed and managed. The exploitation of space in auditory displays has also been considered in recent research. For instance, an interface extending the concept of window systems to audio has been proposed by Ludwig *et al.* [1]. In this example, information is presented with spatially separated sound items that can be monitored and manipulated by the user. The sound sources are spatialised by using Head-Related Transfer Function (HRTF) synthesis and are reproduced over headphones.

3-D sound has been used in different types of auditory displays where sounds have to be separated in space. Begault studied the advantages of a 3-D audio display for plane cockpits, where the pilot is under heavy visual workload [2,3]. Also, background task monitoring using a spatialised audio progress bar has been proposed by Walker and Brewster [4].

Presentation of multiple audio streams to a listener can be achieved with 3-D sound in the same way. For instance, Dynamic Soundscape, a tool for browsing audio data uses the principle of motion of sound sources around the listener [5]. In a similar application called AudioStreamer [6], three audio streams are presented with headphones to the listener at a fixed position (at the front, 60° to the left and 60° to the right, in the horizontal plane). Using a head-tracking technique, this system can get the attention of the user focused on one audio stream by increasing the gain of the source toward which the head is pointing.

Listener's ability to segregate an audio stream in a multi-source environment is termed as the "the cocktail party effect" and was first investigated by C. Cherry in 1953 [7]. In this paper, spatial separation of sound sources is mentioned as an important cue for solving the cocktail party problem. However, as W. Yost mentioned in [8], other cues may also contribute to sound source determination. They include physical attributes of sound such as spectral and temporal content, harmonicity of the signals or temporal onsets and offsets.

The problem of sound source determination and segregation has been widely investigated for speech. For instance, recent research studies compared speech intelligibility and localisation of speech signals in different auditory presentations (monaural/binaural/3-D) [9] or in real and virtual sound-field listening conditions [10]. However, little is known about listeners' ability to localise short non-speech sounds in space, such as earcons and auditory icons. Pitt and Edwards [11] proposed an audio pointer for blind computer users, and considered the identification of speech and non-speech signals with mono and stereo loudspeaker reproduction. In the present study, we focus on non-speech sounds, which are commonly used in auditory displays [12]. A series of tests was performed in order to answer several questions regarding headphone presentation of sound items that are spatially separated. Absolute and relative sound position discrimination is considered in this paper. The effect of the number of items and the temporal overlapping in item presentation were investigated. Also, the differences in performance and preference between these different conditions are discussed.

2. HEADPHONE PRESENTATION USING STEREO PANNING OR 3-D AUDIO PROCESSING

When using headphones, localisation of stereo sound images is usually limited to the lateralisation effect, i.e., sounds are heard inside the head, along the interaural axis. Spatial separation of sources is therefore restricted to this axis and results in unnatural sound perception, especially for signals played only on one channel, i.e., at extreme left or right.

In the case of 3-D sound, a monophonic sound processed with a pair of HRTFs contains the natural cues necessary for sound localisation. When played on headphones, sound is therefore heard outside the head and can be placed virtually at any position around the listener. However, localisation errors are common when non-individual HRTFs are used. These errors include front-back confusions, where sounds are perceived at the reversed position across the frontal plane as described by Carlile [13] and elevation problems for sources presented in the frontal hemisphere as reported by Begault [14]. Also, sound sources presented around the median plane are not properly externalised.

Distance perception can be improved when room reflections are included in the 3-D audio processing. The use of Binaural Room Impulse Responses (BRIRs), which are the equivalent of HRTFs but including also the reflections of the room, increases the proportion of externalised distance judgement, as studied in [15] and [16].

Studies on spatial sound separation usually compare diotic (identical signals at each ear), dichotic (independent signals at each ear) and spatial presentations with HRTFs or BRIRs (individual or non-individual) [9,10]. In the present study, the comparison concerns stereo (i.e. amplitude panning) and non-individual HRTF presentation.

3. SUBJECTIVE TEST DESCRIPTION

Two subjective tests were performed in this study. First, we investigated subjects' ability to spatially discriminate a single sound between five different positions for stereo and HRTF presentation techniques. In the second test, we considered discrimination of a target sound among multiple-position sounds. The three test parameters included in this part of the experiment were:

- The number of sound items presented: 3 and 5 items.
- The processing technique used for sound presentation: stereo and HRTF.
- The onset interval between sound items presented in the same stimulus: 4 different onsets, from successive to simultaneous replay.

A 5-point scale questionnaire was also included in both tests to assess the sound quality regarding the workload for the different replay techniques and the quality of the sound presentation. The following measures were included in the questionnaire: performance, effort, frustration and pleasantness of the sound presentation.

3.1. Stimuli and task

The non-speech sounds used for the tests were melody excerpts of two seconds played with musical instruments of different instrument families generated from the MIDI format. These sounds were chosen because they are easy to recognise and to distinguish from one another. Loudness alignment of the five melody excerpts was performed on the monophonic samples prior to the stereo and HRTF processing. Samples were then created to obtain five different spatial positions. In the case of stereo, lateralisation was achieved by amplitude panning (with a perceptual position adjustment using the equal amplitude technique). For the HRTF technique, 256-tap FIR filters were used with non-individual HRTFs. Azimuth angles chosen were 270º, 320º, 0º, 40º and 90º, all at 0º elevation.

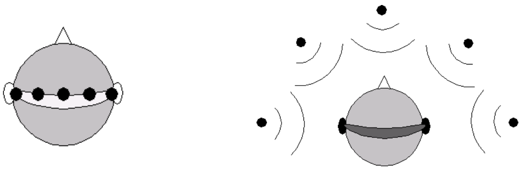

Figure 1. 'Idealised' spatial sound perception of the five positions with headphones for stereo (left) and HRTF presentation (right).

In both tests, we asked the subjects to find the spatial position of a given instrument (flute, guitar, piano, trumpet or violin) between five specific positions; the name of the target instrument was shown on the computer screen prior to the stimulus presentation. Despite the different spatial sound perception expected from the two processing techniques, as ideally illustrated in Fig 1, the same paradigm was used for both techniques in terms of position mapping, i.e. left, between left and centre, centre, between centre and right or right. Also, subjects were familiarised with the selection of these positions in the training phase.

The test subject had to trigger the sound replay from the computer keyboard. In the first test, only one sound item was played at a time. In the second test, three or five sound items were played in a spatial and temporal random order. The subject had to select the position where the target sound item emerged. One of the five keyboard buttons corresponding to the sound positions was used for the selection. An additional button was included to notify a failure to locate the sound item. Quantitative measures of performance included error-rate and response time, i.e. the time between the target sound onset and the position keypress.

Fifteen non-trained listeners, 8 males and 7 females, participated to this experiment. All tests were performed in a quiet room using Sennheiser HD580 headphones. A stimulus delivery program (Presentation Software, from Neurobehavioral Systems Inc.) was employed for the test creation and presentation.

3.2. Test design

The whole experiment was organised in series of two blocks, each block employing one of the two processing techniques. This procedure allowed us to assess each technique separately with the questionnaire. Also, to control the possible habituation and learning effects, the order of presentation of the processing techniques was inverted; half of the subjects received stereo sounds first and other half received HRTF sounds first.

The subjects were first introduced to the concept of locating sounds by making them perform a number of test trials before each test, so that they were familiarised with the different samples and the test procedure.

In the first test, a block of 25 stimuli (i.e. a combination of 5 instruments and 5 positions) was presented to the listener in a random order for each technique. A training session was included in which feedback on success of localisation was given to subject after each answer. The training session was also organised in two blocks in a way similar to the real test. Feedback on localisation success was not given to subjects apart from the training session. This was to stabilise the subjects' possible learning by feedback.

In the second test, the three main parameters included were the number of sound items, the processing technique condition and the onset interval condition (2, 1, 0.5, and 0 seconds). Parameters such as the target sound instrument, spatial position and temporal position, which should have an effect on the subject responses, were not considered in the test design. The same paradigm for Stereo/HRTF block pairs with in-block randomisation was employed but for the two conditions 'number of items' and 'onset interval' (OI), increasing complexity was used for the different block pairs. Each subject started with 3-items, 2s onset interval (sounds played one after the other) down to 3-items, 0 s onset interval (sounds played all together). Then, 5-items conditions were presented, starting with 2s OI down to 0s OI. This summed to 2x4 = 8 scenarios with which the two rendering conditions (stereo/HRTF) were compared.

To take into account the possible effect of differences in instrument and spatial or temporal position, two randomised blocks of 10 cases each were designed to cover a representative sample of all the combinations. An example of spatial and temporal organisation of the sound items for the 5-item condition is shown in table 1. All possible combinations would have equalled up to 125 (5x5x5). This repeated by the number of scenarios (8) would have given us 1000 test stimuli, which was considered not practical for the scope of this study. Using this block design, subjects had to perform only 20x8 = 160 tasks each. The same block of 10 stimuli was used for each processing technique to avoid bias (due to temporal position differences for instance). It was assumed that expectation due to repetition of the same stimuli for each block was not a problem due to the randomisation inside the blocks. Also, different blocks were used for successive scenarios (block 1 for 2s OI, block 2 for 1s OI…).

spatial position (target item in bold)					**temporal position of target**
left	*mid-left*	*centre*	*mid-right*	*right*	
piano	trumpet	guitar	**flute**	violin	5
guitar	flute	**violin**	piano	trumpet	2
trumpet	violin	piano	flute	**guitar**	3
trumpet	piano	flute	guitar	trumpet	1
flute	trumpet	**guitar**	violin	piano	3

Table 1. Examples of spatial and temporal arrangement of the sound items used for the 5-item stimuli.

4. RESULTS OF THE ONE-ITEM TEST

In this section, results regarding the absolute sound position discrimination task are reported. This includes differences between the two processing techniques and the effect of instrument and spatial position. Prior to the test analysis, two assumptions regarding the test design were considered. First, learning effect between the two successive blocks was checked, by looking at the evolution of incorrect responses over the 50 stimuli. For each answer (going from the first stimulus heard to the last one), a mean error-rate was computed over the 15 subjects. No enhancement in performance was observed, Spearman correlation, of value -.002, confirms that error rate remained constant throughout the test.

Then, the order of block presentation, i.e. Stereo/HRTF for group A and HRTF/stereo for group B, were compared. No effect on subjects' performance was found. Difference between the two groups is not significant (2-tail t-test, equal variances assumed: t=0.869, df=118). However, differences in response times were observed between the two groups. Due to instruction problems, large variations in response times were found within Group A. So, it was decided to include group B only in the response time analysis.

4.1. Comparison of the two presentation techniques

The amount of incorrect responses, i.e. keypress answers that do not match with the target position were first compared for the two presentation techniques. The total count of responses was 750 (25x2 techniques by 15 subjects) and none of the listeners used the '*can't localise*' answer in this part of the test. A significant difference between the two presentation techniques (see table 2) was found in favour of the stereo presentation (*t= -2.845, df=748, p=.005*). Due to large variations in error-rates between subjects (from 0 to 56% errors for 25 stimuli), differences in error-rate between the two techniques were also computed for each subject. A difference of 9.6% was found in favour or stereo.

proc. tech.	**Error-rates (%)**	**SD (%)**
Stereo	14.9	10.5
HRTF	22.7	12.2
Diff(H-S)	9.6	6.0

Table 2. Comparison of error-rates for the two presentation techniques,

The two presentation techniques were then compared in terms of response time (RT), i.e. the time between the target sound onset and the position keypress. Only RTs for correct responses were included in the analysis. Also, due to the difference between the two groups mentioned in 4.1, only the group B was considered for RT analysis (324 cases, as shown in Table 3). No significant difference was found between the two conditions (2-tail t-test). As identical stimuli were tested for the two presentation techniques, the difference in RT for identical stimulus was also computed. A mean value of 4 ms only was found for the difference.

proc. tech.	**cases**	**mean (ms)**	**SD**
Stereo	167	1240.0	624
HRTF	157	1266.1	706
Diff(H-S)	135	-3.9	565

Table 3. Comparison of response times (in milliseconds) for the two presentation techniques.

Finally, qualitative gradings of the workload and pleasantness of the sound presentation were compared for stereo and HRTF. A significant difference was found for two of the aspects included in the questionnaire in favour of stereo, namely *performance* (*2-tail t= -2.869, df=14, p<.05*) and *effort* (*2-tail t = 4.000, df=14, p .01*). It can also be noticed that a negative correlation of -0.58 was found between the two gradings, which explains the inverse relation between these two scales (Table 4).

Question	**Stereo**	**HRTF**	**Dif(H.,Ste.)**	**ttest**
Performance	3.67	3.00	**-0.68**	0.01
Effort	2.40	3.20	**0.80**	0.00
Frustration	4.01	3.54	-0.47	0.12
Sound quality	4.32	3.87	-0.47	0.10

Table 4. Qualitative assessment compared for the two presentation techniques. Differences are computed for each subject.

4.2. Effect of instrument and spatial position

The two other factors of interest in this first test were the instruments and the spatial position of the target sound items. Each instrument was presented 150 times (5 positions by 2 techniques for 15 subjects). The guitar sound was localised significantly worse (t-test t=2.81, df=297, p<.01) than the piano sound (26% and 13% error-rate respectively). The same error-rate, 18%, was found for the three other instruments. However, looking at the two presentation techniques separately, more errors were observed for HRTF than stereo, in the case of the piano, trumpet and violin sounds (~10% and ~20% error-rate respectively).

Spatial position had also a significant effect on the amount of incorrect responses (*chi-square=20.771, df=4, p<.001*). Due to the differences in spatial perception for stereo and HRTF presentations, error rates were considered for each presentation technique separately, as depicted in Fig. 2. Overall, a similar trend is observed amongst positions for the two techniques, with significantly better localisation for the centre and right positions (t-tests between positions: centre and right differ from others with *p<.05*). Unexpected asymmetry in incorrect response percentage was also observed between left and right positions.

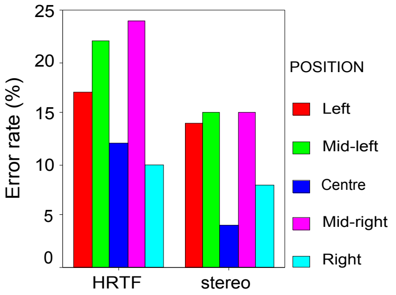

Figure 2. Error-rate per spatial position as a function of the presentation technique for the one-item test.

4.3. Summary of results for the one-item test

In this first test, listeners' ability for absolute sound position discrimination was considered by presenting one sound item at five different positions. Performance in this task was measured by error counts; an error-rate of 18.8% was found for all stimuli. Looking at the error distance between target position and response, we observed a very small amount of errors with a distance superior to 2 (less than 1%). This gives an indication that a high localisation performance would be achieved in a similar task with only the three positions: left, centre and right.

Comparing the two presentation techniques, stereo performance was found significantly better than the HRTF one. This technique was also graded higher on the *performance* and *effort* scale in the qualitative questionnaire. However, no significant difference was observed in response time between the two techniques. Differences in error-rates between instruments (especially *piano* and *guitar* sounds) and spatial positions were also observed in this one-item test.

5. RESULTS OF THE MULTIPLE-ITEM TEST

Results regarding the relative sound position discrimination task are reported in this section. Differences between the two processing techniques are considered first. Then, the other effective factors are analysed, including the number of items, the onset interval, the spatial position and temporal position of the sounds.

Taking into account results from the first test, we assumed that learning effect was not significant within series of two 2 blocks. Indeed, only 20 stimuli were included in each series (10 per block), in comparison to the 50 stimuli used in the first test. Assumption on the effect of the group order was also checked. No significant difference between group A (Stereo first) and group B (HRTF first) was observed. (2-tail t-test, equal variances assumed: *t=-1.056, df=2398*). Finally, only responses of the group B were included in the response time analysis for the same reason as in the first test.

5.1. Presentation technique effect

From the response data, the amount of successful localisation and response time to correct target stimuli were compared amongst HRTF and stereo. Total response count was 1200 per technique (2x80 trials per subject). The overall error-rates are 21.8% for stereo and 23.8% for HRTF, but the difference is not statistically significant (chi-square = 1.482, df=2). 'Can't localise' answers, which are included in the incorrect responses, counted for 4.8% and 4.7% for stereo and HRTF respectively.

Response times were next turned to, to look for possible difference between processing techniques. Again, response times are also limited to correct responses (1017 cases). Average response times (in milliseconds) and error-rates are depicted in Table 5, per item as a function of OI and processing technique condition. None of the differences between processing techniques are significant in any condition (2-tail t-tests).

		3 items			5 items		
OI	**tec.**	e-r (%)	rt (ms)	cases	e-r (%)	rt (ms)	cases
2 s	**stereo**	9.4	1600	77	24.7	1685	65
	HRTF	12.7	1800	71	22.7	1813	62
1 s	**stereo**	8.7	1363	76	20.0	1620	68
	HRTF	8.0	1341	75	30.7	1557	59
0.5 s	**stereo**	6.0	1329	75	30.0	1616	58
	HRTF	10.7	1324	73	36.0	1473	51
0 s	**stereo**	16.7	1777	68	58.7	2100	36
	HRTF	14.7	1922	70	55.4	2048	33

Table 5. Comparison of error-rates and response times (in milliseconds) as a function of item number, presentation technique and onset interval.

5.2. Other effective factors

5.2.1. Number and onset interval of the sounds

The number of the sound items and the onset interval had an important effect on localisation performance. A univariate linear model was calculated, using localisation errors per block of 10 stimuli as the dependent variable, and onset interval and number of items as independent variables. The model explains 66.4% of the variance in response errors (*adjusted R^2 = 0.664*, with all the variable and intercept effects being significant at *p<.001*).

A mean error was computed per block of 10 stimuli over all subjects. As shown in Table 6, errors rise as the onset interval shortens and number of items increases. Differences between sound item conditions are significant as whole (*1-way ANOVA with df=1, F=83.399, p<.001*) and differences between the sound onset interval conditions are also significant (*1-way ANOVA with df=3, F=8.678, p<.001*).

3 items				5 items			
OI	N	er.rate (%)	SD	OI	N	er.rate (%)	SD
2 s	15	11.0	9.9	**2 s**	15	23.7	11.4
1 s	15	8.3	12.5	**1 s**	15	25.3	9.7
0.5 s	15	8.3	6.5	**0.5 s**	15	33.0	12.9
0 s	15	15.7	9.8	**0 s**	15	57.0	12.4
Total		10.8		Total		34.5	

Table 6. Mean error-rates, per 10 stimuli, as a function of onset interval for the 3-item and 5-item conditions.

Considering now response times as a function of OI within each item condition, 1-way ANOVA tests indicate that response time differences between the conditions are significant (*df*s=*3* and *1, F*s=*19.677* and *8.400*, *p* values *<.001* and *.005*, respectively). There appears nonlinearity in effect of the onset interval: the highest response times occur both at 2 s interval (successive presentation of sounds) and 0 s (simultaneous presentation). Average response times and 95% confidence intervals are depicted in Fig. 3. The response times can be verified separately for each processing technique from table 5.

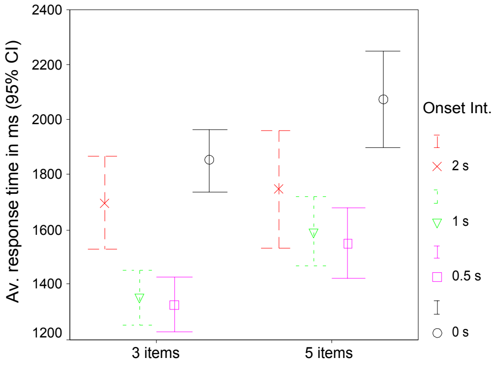

Figure 3. Means and 95% confidence interval for the response time in the multiple-item test as a function of number of items and onset interval.

5.2.2. *Spatial position of target sound*

The spatial position of target sound had also an effect on localisation performance. Table 7 depicts the error-rate in 3- and 5-item conditions. With 3 items, the difference against centre position is not significant (*chi-square=14.946, df=8*). For 5 items, the difference against mid positions and in favour of centre position is significant (*chi-square=94.135, df=8, p<.001*). The same left/right asymmetry is observed for the 5-item condition, as in the first test.

Item cond.	***spatial position of target sound***				
3 items	*left*		*centre*		*right*
incorrect	8.9%		15.6%		9.0%
5 items	*left*	*mid-left*	*centre*	*mid-right*	*right*
incorrect	47.2%	52.9%	18.7%	39.6%	25.0%

Table 7. Error-rate per spatial position for 3 and 5 items.

5.2.3. *Temporal order of target sound*

Table 8 depicts the differences in error-rates in relation to the temporal order of target sound, by item condition. It can be seen, that when the target was the first sound presented, its position was less accurately localised. Difference is significant for both 3-item (*chi-square=11.664, df=2, p<.05*) and 5-item conditions (*chi-square=133.356, df=2, p<.01*).

number of items	***temporal order of target sound***				
	1	**2**	**3**	**4**	**5**
3 items, incorrect	12.7%	7.5%	10.0%	-	-
5 items, incorrect	51.3%	36.2%	18.4%	26.9%	26.8%

Table 8. Error-rate per temporal position for 3 and 5 items.

5.3. Qualitative assessment

Regarding the qualitative gradings of the workload and pleasantness of the sound presentation, no differences were observed between stereo and HRTF presentation techniques. Assessment results between 3- and 5-item conditions show significant differences in all four aspects (*performance*, *effort*, *frustration* and *sound quality*) in favor of 3 items (1-way ANOVA, df=1, p's <.05).

Assessment results compared between the onset interval conditions also show differences between the four conditions (1-way ANOVA, df=3, p's <.01). In 0-second OI cases, performance (F=15.361) and sound quality (F=4.001) are assessed lower and effort (F=7.724) higher than in other cases. Interestingly, frustration (F=3.953) is assessed lower also. Figure 4 depicts the assessment value averages (with 95% confidence intervals) by onset interval.

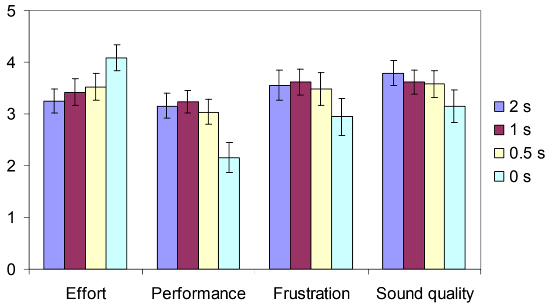

Figure 4. Means and 95% confidence interval for the preference grading in the multiple-item test, as a function of the onset interval.

5.4. Summary of results for the multiple-item test

In this second part of the study, listeners' ability to discriminate spatially one sound item in the presence of competing sounds was tested. The two important factors observed in this experiment were the item number (3 or 5) and the onset interval (0, 0.5, 1 or 2 s). However, no significant difference was found for the presentation technique condition (stereo or HRTF) in terms of error rates, response times and qualitative aspects.

Considering incorrect responses as a function of the item number condition, an error-rate of 10.8% in 3-item condition, and 34.5% in 5-item condition were observed. Also, in the 3-item condition, onset intervals 0.5s and 1s are significantly lower in terms of error-rates (8.3%) and response times (~1300 ms).

As in the first test, large differences in error-rates between spatial positions were observed. In the 5-item condition, error rates are significantly lower for the centre position and higher for mid-positions. It was also noticed that the temporal order has an effect, with a large error rate for target sounds being presented first.

6. DISCUSSION

The absolute spatial discrimination task of the first test showed that non-trained listeners are able to distinguish between 5 positions to a moderate degree of accuracy with both the stereo and HRTF presentation techniques. Results also indicate that using only 3 positions would considerably decrease the number of errors in this task. The higher performance and preference observed for stereo may be taken as a surprising result, considering the numerous research works on the advantages of HRTF presentation for sound localisation. However, a direct comparison with existing results is difficult because literature on headphone studies including stereo panning and HRTF technique is not extensive. Also, studies on the cocktail party phenomenon usually consider speech intelligibility and speaker-recognition, and compare monaural, binaural and 3-D audio presentations [9]. In the present study, we proved that stereo panning allows an efficient spatial separation for a limited number of sound items. The use of non-speech sounds also makes a difference with existing literature. In Pitt and Edwards' study [11] on auditory selection between multiple sounds using stereo panning over loudspeakers, differences between voices and non-speech (musical) sounds were also observed. Finally, we should mention that preference for stereo is maybe attributed to the lack of familiarity with the presentation techniques for the non-trained listeners employed in this test. Long-term exposure to the sounds would certainly give different results, due to the fatigue caused by the inside-the-head perception in stereo presentation.

Results observed in the second test proved that multiple-item presentation is feasible with both presentation techniques. This gives credit to HRTFs as an efficient technique for spatial separation of sound sources. With more than 30% error-rate, the 5-item condition may be seen as critical from the usability viewpoint, but it should be noted that the random order used for sound item presentation (rather than playing sounds from left to right) is quite artificial and makes the localisation task a lot more difficult. Accuracy in the multiple-item task depends on the number of items, but the onset time between sound items has also a remarkable effect on response time and error-rate; a short onset time (0.5s) make a significant difference in localisation accuracy compared to simultaneous replay.

Considering spatial positions now, a problem with mid-positions was observed in both tests. In the case of stereo panning, this could be due to a lack of spatial separation from the centre position. With HRTFs, source images may not be clearly defined for the 320º and 40º azimuth positions, due to front-back confusion, elevation problems or lack of externalisation. A combination of stereo and HRTF, employing 270º and 90º azimuth HRTFs and stereo panning for other positions would maybe work better. Also, the differences in performance observed between sound items for absolute and relative localisation is certainly due to the frequency content of the sounds. This affected the sound spatialisation and produced masking between sounds.

Finally, considering the item condition for both tests, we see a difference between 1 or 3-item vs. the 5-item condition rather than single vs. multiple-item condition. This proves that presenting three sounds does not cause problems for the user.

7. CONCLUSIONS

In this paper, a series of experiments was carried out to study spatial position discrimination with a limited number of non-speech sounds. This test showed that presenting multiple sound items over headphones is feasible to a certain extent. Stereo panning and HRTF processing were compared for one-item, 3-item and 5-item presentation. Within the restricted conditions of the present test, differences between the two presentation techniques were observed in favour of stereo for the one-item condition; but no significant difference was found in the multiple-item test. An optimal performance was found in the second test for the 3-item condition with 1 s and 0.5 s onset intervals. In the future, tests involving user interaction in this type of simple spatial auditory display should be considered.

8. REFERENCES

[1] L. Ludwig, N. Pincever and M. Cohen, "Extending the Notion of a Window System to Audio", IEEE Computer, August 1990, pp. 66.

[2] D. Begault, "Head-Up auditory display research at NASA AMES Research Center", in Proc. of the Human Factors and Ergonomics Society 39th annual meeting, pp. 114-118. Santa Monica, California: Human Factors Society, 1995.

[3] D. Begault, E. Wenzel, R. Shrum and J. Miller, "A virtual Audio Guidance and Alert System for Commercial Aircraft Operations", in Proc. Of ICAD'96, Palo Alto, California, November 1996.

[4] A. Walker and S. A. Brewster, "Spatial audio in small display screen devices", Personal Technologies, 4(2), pp 144-154 (2000).

[5] M. Kobayashi and C. Schmandt, "Dynamic Soundscape: mapping time to space for audio browsing", in Proc. CHI'97, ACM Press Addison-Wesley, 1997.

[6] C. Schmandt and A. Mullins, "AudioStreamer: Exploiting Simultaneity for Listening", in Proc. CHI'95. ACM Press Addison-Wesley, 1995.

[7] C. Cherry, "Some experiments on the recognition of speech, with one and two ears", J. Acoust. Soc. Am., vol. 26, pp. 975 (1954).

[8] W. A. Jost, "The Cocktail Party problem: Forty Years Later" in "Binaural and Spatial Hearing in real and Virtual Environments", R. H. Gilkey, T. R. Anderson (eds.). Erlbaum, NJ, pp. 329-347, 1997.

[9] R. Drullman and A.W. Bronkhorst, "Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimentional auditory presentation", J. Acoust. Soc. Am., vol. 107, pp. 2224 (2000).

[10] M. L. Hawley, R. Y. Litovsky, and H. S. Colburn, "Speech intelligibility and localization in a multi-source environment", J. Acoust. Soc. Am., vol. 105, pp. 3436 (1999).

[11] I. J. Pitt and D. N. Edwards, "Pointing in an Auditory Interface for Blind Users", in Proc. of the 1995 IEEE Conference on Systems, Man and Cybernetics, IEEE, 1995.

[12] S. A. Brewster, V.-P. Raty and A. Kortekangas, "Earcons as a Method of Providing Navigational Cues in a Menu Hierarchy", in Proc. of HCI'96 (Imperial College, London, UK), Springer, pp 167-183 (1996).

[13] S. Carlile, "Virtual auditory space: Generation and applications" Austin, Landes: Ch. 1, 1996.

[14] D. R. Begault, "Perceptual similarity of measured and synthetic HRTF filtered speech stimuli", J. Acoust. Soc. Am., vol. 92, pp. 2334 (1992).

[15] D. R. Begault, E. M. Wenzel, A. S. Lee and M.R. Anderson, "Direct comparison of the impact of Head Tracking, Reverberation, and Individualized HRTFs on the Spatial Perception of a virtual Speech Source", Audio Eng. Soc. 108th Convention, Paris, France, February 2000, preprint no. 5134.

[16] G. Lorho, "Virtual Source Imaging System Using Headphones", M.Sc. Thesis, December 1998, ISVR, Univ. of Southampton, UK.