DO LOCATION AND CONTENT OPERATE INDEPENDENTLY?

Paul M. C. Lemmens, Ab de Haan, Gerard P. van Galen

University of Nijmegen
P.O. Box 9104, NL–6500 HE Nijmegen, The Netherlands
P.Lemmens@nici.kun.nl

ABSTRACT

To investigate if location and content operate independently, two experiments were carried out in which redundant auditory co–messages to a visual categorization task were manipulated to enable left and right ear only presentation, as well as binaural presentation. The location cues were irrelevant to the main task and the auditory co–messages (in one experiment earcons and in the other auditory icons) contained redundant information from the primary task. The findings seem to indicate that location and content of the co–messages operate independently because no interaction between these factors were found.

1. INTRODUCTION

3D presentation of auditory scenes is becoming increasingly popular in home-theaters, movie theaters, and computer environments, both lifelike (using multiple speakers, e.g., [1]) as well as virtual (using well known auditory parameters that can trick a user in perceiving a 3D sound environment with a limited amount of speakers, e.g., [2]). In a movie theater, seeing a gun being fired at your (front) right, hearing the bullet whistle by impacting somewhere behind you at your rear left, will pull your perceptual space [3] into the movie, engaging it more actively than in a simple stereophonic movie. Active engagement will increase the sense of presence [4] dramatically, thus increasing the entertainment value [5].

Most experiments investigating effective 3D sound presentation, manipulate some parameters in the sounds that are thought to be related to 3D perception of sound. Usually the participant's task is to track or point to the perceived location (e.g., [1, 6]), the sound itself being white noise or simple beeps. Such experiments have pointed out many parameters relevant for 3D perception of sound. However, in settings in which the pure location of a sound is of lesser importance (or even irrelevant), and the content (e.g., as co–message to a visual task) is of higher importance, the location cue can interact with content in unexpected ways, by overlap of the location cue with another stimulus feature.

One such unexpected interaction is a finding in experimental psychology known as the Simon effect, that builds upon the interaction between an irrelevant location component of a stimulus with an irrelevant location component of the required response (for a review by Simon see [7]). For instance, if a task requires participants to press a red button that is located to their left if they see a red dot, the presentation of the red dot to the right of the centerline of a screen increases response times (RT) significantly compared to RT's to the same red dot presented on the left of the centerline. In this example, location and content do not operate independent, because the irrelevant location of the stimulus affects how effective participants respond to the relevant content (red vs. green). The differences between the corresponding (left–left) and noncorresponding (left–right) conditions can reach up to 30 to 40 ms and survive many experimental manipulations, including responding with crossed hands. Originally, the phenomenon was explained as a natural tendency to respond towards the location of stimulation. Although this simple description still has some viability, to date findings exist that need a more complex explanation than just a natural tendency.

In our earlier experiments [8, 9, 10, 11] we have consistently found an effect, comparable to the Simon effect, of content of an auditory co–message (an auditory icon or an earcon) to a categorization task of pictures of animals and musical instruments. For both the auditory icons and the earcons, congruent combinations (e.g., animal picture and animal sound) were processed more effectively than incongruent combinations (animal picture and instrument sound). For the earcons a congruent picture–sound combination always was either the picture of an animal with a major chord or the picture of another object with a minor chord. Incongruence was congruence mirrored, for instance, pictures of animals with a minor chord and pictures of other objects with a major chord. In general, the experiments showed that auditory icons (usually sounds of animals and musical instruments) as flankers[1] facilitate response time latencies to the categorization task (Exp. 2 in [10] and Ch. 5 in [8]) whereas earcons (major and minor chords) slow down re-

[1]The earlier experiments termed the auditory stimuli distracters. Because these sounds were never intended to distract a participant, we now call them flankers, in the tradition of the Eriksen flanker task, and also sometimes co–messages.

sponses [9, 12].

These earlier experiments have clearly shown that the cognitive information processing system is capable of creating a unified percept out of these multi–modal stimuli, much like, for example, the visual and auditory streams of a ventriloquist's performance are integrated into a compound percept. However, how do the individual features of such multimodal stimuli interact? For instance, will the effect of the content of the auditory part of the stimulus be different if it is presented in different (spatial) locations? Two experiments were conducted using auditory icons and earcons as auditory flankers, in adapted versions that could be presented to the left ear exclusively, to the right ear exclusively, and (in the regular way) to both ears simultaneously. Because participants had to respond bimanually using response buttons on either side of their visual fixation point, a Simon task was created as the presentation location of the auditory flanker could overlap with the location of the (required and correct) response button.

2. GENERAL METHOD

Experiment A employed auditory icons that could be presented to the left ear only, to the right ear only, and binaurally (stereo). Experiment B used earcons, in the same presentational variations.

The auditory icons in Experiment A were two musical sounds (a piano playing and the sound of a drum) and two animal sounds (a dog barking and a chicken cackling). All sounds had a duration of a little over 1200 ms and were presented simultaneously with the pictures consisting of seven animals and seven musical instruments. The pictures were black and white line–drawings. The earcons in Experiment B were the major and minor version of the triad at C5, with a duration of 2500 ms. Instead of the category of musical instruments, Experiment B used objects other than musical instruments as non–target category.

A randomized within–subjects design was constructed using two factors. The correspondence between the presentation location of the auditory flanker and the location of the (correct) response button, the Simon task (called Spatial Correspondence), had three different conditions (corresponding, noncorresponding, and stereo). The corresponding condition included all trials in which the correct response location corresponded with the presentation location of the auditory flanker (left ear or right ear); for the noncorresponding condition this relation was reversed. The stereo condition comprised all trials in which the auditory flankers were presented binaurally. The mapping between picture and sound was an experimental factor with two levels: congruent and incongruent. For Experiment A the congruent Picture–Sound Mapping consisted of those trials in which animal pictures were accompanied by a sound of their own category (e.g., the picture of a bird with the cackling of a chicken); for Experiment B the appropriate sound was the major chord. The incongruent condition reversed this relation. Baseline trials were added by randomly intermixing all experimental trials with an equal number of trials without sound. These trials were named the silent (or no–sound) condition.

The experiments were carried out on a Macintosh G3 computer that was equipped with a buttonbox for timing the stimulus presentations and registering response times. The auditory flankers were presented over simple stereophonic headphones. Experiment A was conducted with 20 participants; Experiment B used 18 participants. Most participants (for both experiments, mean age approx. 24 years) were students in Psychology or Cognitive Science at the University of Nijmegen. The participants either received money or course credits for their participation.

3. RESULTS

Before any statistical tests were carried out the data of both experiments were pruned from incorrect responses and response omissions. This procedure amounted to the removal of about 2.4% of the total number of data points per experiment.

All statistical tests were carried out using a multivariate repeated measures ANOVA for the main effects and univariate anova's for contrasting individual conditions, using a significance level of $\alpha = .05$.

3.1. Experiment A: left/right auditory icons

Auditory flankers facilitated response times (mean RT 416 ms) compared to the baseline (423 ms) and tended towards significance: $F(1, 19) = 4.069$, $p = .058$. Such a general facilitation (or inhibition) of RT's is termed an effect of sound.

Spatial Correspondence showed a significant main effect ($F(2, 18) = 17.572$, $p < .001$) with a significant difference between the corresponding (403 ms) and noncorresponding condition (428 ms; $F(1, 18) = 37.870$, $p < .001$).

There also was a main effect for Picture–Sound Mapping, indicating that the difference between the congruent condition (413 ms) and the incongruent condition (421 ms) was significant: $F(1, 19) = 12.004$, $p < .05$.

No interaction ($F < 1$) was found for Picture–Sound Mapping and Spatial Correspondence. The lines in Figure 1 (left panel) are nearly parallel, visually confirming the absence of an interaction. Table 1 presents mean response time latencies for the levels of the factorial combination of Picture–Sound Mapping and Spatial Correspondence in this experiment.

Table 1: *Mean response time latencies (ms) for all conditions of Picture–Sound Mapping and Spatial Correspondence for both experiments. Experiment A used auditory icons; Experiment B used earcons. In braces is the standard error.*

	Experiment A		Experiment B	
	Picture–Sound Mapping			
Spatial Correspondence	**Congruent**	**Incongruent**	**Congruent**	**Incongruent**
Corresponding	400 (10)	405 (11)	403 (11)	407 (11)
Noncorresponding	425 (11)	431 (10)	425 (11)	427 (11)
Stereo	413 (10)	427 (11)	411 (12)	415 (11)

3.2. Experiment B: left/right earcons

In this experiment using earcons with a left/right location cue, no general effect of sound was found (with sound 415 ms, without sound 415 ms; $F < 1$). There also was no effect of Picture–Sound Mapping (congruent 413 ms, incongruent 417 ms; $F(1,17) = 1.31$, $p > .25$).

Spatial Correspondence, however, showed a very strong main effect ($F(2,16) = 7.113$, $p < .05$) and, most importantly, a significant difference between the corresponding condition (405 ms) and the noncorresponding condition (413 ms; $F(1,17) = 14.114$, $p = .002$).

In this experiment also, there was no interaction between Picture–Sound Mapping and Spatial Correspondence ($F < 1$). Figure 1 (right panel) again shows almost parallel lines. Table 1 also shows the mean RT's for Experiment B.

4. DISCUSSION

As in earlier experiments that employed auditory icons as flankers (e.g., Ch. 4 & 5 in [8], and [10]), Experiment A showed faster response times to congruent Picture–Sound Mappings than to incongruent ones. Unfortunately, there was no effect of Picture–Sound Mapping in Experiment B (using earcons). This may have been caused by the randomized design in Experiment B. Earlier experiments that employed earcons in a randomized design have also shown diminished effects of Picture–Sound Mapping, or even the complete absence of such an effect, compared to experiments employing a blocked design (cf., Exp. 1 and Exp. 3 in [13]).

Both experiments, however, clearly show the anticipated Simon effect. If the presentation location of the auditory flanker did not correspond with the location of the required button, responses were significantly slower, whereas a corresponding location facilitated responses (both compared to the no–sound baseline trials). These findings are another corroboration of earlier findings that the (original) Simon effect can occur in multi–modal settings [7, 14] and imply that people can make accurate left/right discriminations and that they use this information to their advantage if the location cue is be informative with respect to the task that is being carried out (despite that the location cue actually is irrelevant for the main task).

The important finding in both experiments, however, is that there was no interaction between Picture–Sound Mapping and Spatial Correspondence, as Figure 1 shows. The lines are almost parallel, meaning that for both experiments, for each condition of Spatial Correspondence, the differences between congruent and incongruent Picture–Sound Mappings were almost the same. Although the stereo condition in Experiment A seems to show an interactive effect, this effect was not significant at all. The expectation that location and content operate independently can therefore be confirmed: in the present experiments location has no detrimental effects on the effects of the content of the auditory flankers. These findings imply that as long as an auditory co–message is irrelevant to the main task, a location component can be added to the co–message that can and will direct attention to another position in space. The shift of attention caused by the location cue has no effect on the co–message itself.

Several examples can be thought up that may indicate the practical relevance of the current findings. For example, consider a task in virtual 3D environment that consists of a number of known sequential tasks. If the user unwittingly starts with the second task instead of the first, an auditory warning signal should facilitate the ease of detecting the failure of adhering to the proper sequence. The current findings make it possible to include a location cue in the warning signal towards the position where the first task should have been carried out. This way, a specific sound that draws attention to another part of the environment can be omitted, resulting in a less confusing auditory environment. Alternatively, on a PDA information is likely to scroll of the screen easily and quickly, for example in a spreadsheet. A location component in the co–message belonging to the current manipulation of data in cell (x, y) can point to the location of another cell who's content will be changed due to the handling of the data in the current cell. These examples both assume that the location component as well as the auditory signal itself are irrelevant to the primary task, that is, both the 3D and the PDA task can be carried out successfully with no auditory input or feedback.

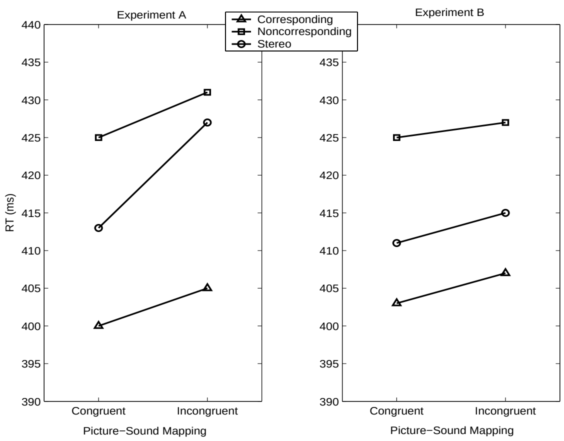

Figure 1: *Mean RT's (ms) for congruence vs. incongruency for each Spatial Correspondence condition for Experiment A (left panel) and Experiment B (right panel).*

The current study shows that location and content in an auditory co–message do operate independently, at least if the location cue is irrelevant to the task at hands. To enable more solid theoretical conclusions or practical guidelines, additional experiments have to be carried out to investigate the existence of the Simon effect in situations where the auditory co–message provides complementary information to the event it accompanies, because it may be that in such a case an interaction between content (Picture–Sound Mapping) and location (Spatial Correspondence) may actually occur. Unfortunately, the earcons in Experiment B did not show any indication of their regularly found inhibition of RT's in other studies. This is also a finding that needs further investigation.

5. REFERENCES

[1] J. A. Ballas, D. Brock, J. Stroup, and H. Fouad, "The effect of auditory rendering on perceived movement: Loudspeaker density and HRTF," in *Proceedings of the 7th International Conference on Auditory Display*, J. Hiipakka, N. Zacharov, and T. Takala, Eds. 2001, pp. 235–238, ICAD.

[2] G. Lorho, J. Marila, and J. Hiipakka, "Feasibility of multiple non–speech sounds presentation using headphones," in *Proceedings of the 7th International Conference on Auditory Display*, J. Hiipakka, N. Zacharov, and T. Takala, Eds. 2001, pp. 32–37, ICAD.

[3] P. P. Lennox, J. M. Vaughan, and T. Myatt, "3D audio as an information–environment: Manipulating perceptual significance for differentiation and pre–selection," in *Proceedings of the 7th International Conference on Auditory Display*, J. Hiipakka, N. Zacharov, and T. Takala, Eds. 2001, pp. 155–160, ICAD.

[4] W. Barfield, D. Zeltzer, T. B. Sheridan, and M. Slater, "Presence and performance within virtual environments," in *Virtual Environments and Advanced Interface Design*, W. Barfield and T. A. Furness, Eds. 1995, Oxford University Press.

[5] N. Lodge, "Being part of the fun – immersive television!," 1999, Keynote address at the Conference of the Broadcast Engineering Society (India, New Delhi), 2–4 february 1999.

[6] E. M. Wenzel, "Effect of increasing system latency on localization of virtual sounds with short and long duration," in *Proceedings of the 7th International Conference on Auditory Display*, J. Hiipakka, N. Zacharov, and T. Takala, Eds. 2001, pp. 185–190, ICAD.

[7] J. R. Simon, "The effects of an irrelevant directional cue on human information processing," in *Stimulus–Response Compatibility*, N. Cowan and A. Barron, Eds., North–Holland, 1990, pp. 31–86, Elsevier Science Publishers.

[8] M. P. Van Esch–Bussemakers, *Do I hear what I see? The Influence of Auditory Icons and Earcons on Multimodal Integration in Visual Categorization*, Ph.D. thesis, NICI, University of Nijmegen, 2001.

[9] P. M. C. Lemmens, M. P. Bussemakers, and A. de Haan, "The effects of earcons on reaction times and error–rates in a dual task vs. a single task experiment," in *Proceedings of the International Conference on Auditory Display*, P. R. Cook, Ed. 2000, pp. 177–183, International Community for Auditory Display.

[10] P. M. C. Lemmens, M. P. Bussemakers, and A. de Haan, "Effects of auditory icons and earcons on visual categorization: The bigger picture," in *Proceedings of the International Conference on Auditory Display*, J. Hiipakka, N. Zacharov, and T. Takala, Eds. 2001, pp. 117–125, International Community for Auditory Display.

[11] P. M. C. Lemmens, A. de Haan, and G. P. van Galen, "Visual categorization with congruent or incongruent auditory flankers including a location component," *Psychological Research*, submitted.

[12] M. P. Bussemakers, A. de Haan, and P. M. C. Lemmens, "The effect of auditory accessory stimuli on picture categorization; implications for interface design," in *Human–Computer Interaction: Ergonomics and User Interfaces, Volume 1 of the Proceedings of the 8th International Conference on Human–Computer Interaction*, H. Bullinger and J. Ziegler, Eds., London, 1999, vol. 1, pp. 436–440, Lawrence Erlbaum Associates.

[13] M. P. Bussemakers and A. de Haan, "Getting in touch with your moods: Using sound in interfaces," *Interacting with Computers*, in press.

[14] J. R. Simon and J. L. Craft, "Communicating directional information with an auditory display," *Journal of Applied Psychology*, vol. 55, no. 3, pp. 241–243, 1971.