EFFECTS OF AUDITORY ICONS AND EARCONS ON VISUAL CATEGORIZATION: THE BIGGER PICTURE

Paul M. C. Lemmens, Myra P. Bussemakers, Abraham de Haan

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

ABSTRACT

This paper presents an overview of work on the effects of earcons and auditory icons on picture categorization and the results of 2 new experiments. The general finding of the experiments is that earcons have an inhibitory effect on picture categorization whereas auditory icons, in general, have a facilitating effect. These findings will be discussed and related to the theoretical framework of perceptual versus conceptual categorization.

1. INTRODUCTION

Imagine dubbing a digital movie of your pet dog and cat chasing each other, with the sound of a flock of quacking ducks. The resulting tape would probably both be funny as well as completely unnatural, thus alerting and attracting attention. Modern computer systems, on the other hand, are equipped with the most peculiar sound schemes one can think of: relating the crowing of a rooster to the arrival of new email is less than trivial. Yet, assigning such nontrivial sounds to events on a computer is common practice, whereas only a few people would (or, in any case, could) train their dog to quack like a duck.

The fundamental difference between these combinations of a sound and a visual event is that the crowing email arrival is used in a signaling context where the objective of the sound is to alert the user to the occurrence of an event. The quacking dog, on the other hand, is situated in a co–messaging context where, in this case, the auditory signal strangely modulates the event, because the visual and auditory signals signify different categories.

In user interfaces, alarm sounds usually provide additional information to the information already present in the visual part of the interface and research investigating the effects of this additional information usually indicates that performance (e.g., measured in product throughput) is enhanced [1, 2]. However, as Edworthy [2] noted, much of the performance enhancements Rauterberg [1] found can be related to the addition of new information, not to the use of the sounds per se.

Research on the effects of sounds carrying *redundant* information is scarce however and this paper provides a summary of the work done at the Nijmegen Institute on Cognition and Information on the effects of auditory icons and earcons, carrying redundant information, on response times to a picture categorization task. The paper also presents two new experiments that have been carried out recently and relates these new findings to the previous results and a theoretical framework.

The notion of sounds carrying *redundant* information is critical to the understanding of the research: The criticism of Edworthy [2], on the conclusions Rauterberg [1] drew, is that the experimental setup Rauterberg used, did not allow for concluding that the addition of sound led to performance enhancements, because the effects of the addition of sound (to which Edworthy attributes the findings of Rauterberg), cannot be discriminated from the effects due to the addition of new information (although presented by sound). In other words, had the information in the sound been presented visually, Rauterberg would have found the same performance enhancements as in his original study. Substituting visual cues by auditory ones does not conform to definition of redundant information in the auditory distracters that Bussemakers and de Haan use. For example, making the buttons on a small user interface smaller and adding non–speech sounds to present information about the buttons, like Brewster [3] does, is different from adding sounds to an interface that are completely redundant and, thus, unnecessary. For equivalent performance, compared to the original button sizes, Brewster needed the sounds to help boost accuracy.

The notion of *redundancy* of information therefore is not a notion of removing information from a stimulus in one modality and presenting that information via an additional stimulus in another modality, nor is it the notion that new information is presented in an additional modality. The definition of redundant information (in a secondary modality), that Bussemakers and de Haan use, is information that is unnecessary to satisfactorily fulfill the primary task at hand. That is, the redundant information presented in the secondary modality is completely unnecessary to respond appropriately to the stimuli in the primary modality at the same level of performance (at the level of response times as well as error–rates) compared to the level of performance without the redundant (auditory) information, Supplying redundant auditory information does provide the experimenter with the possibility of separating the effects of the auditory components of an event and the information this auditory component carries. This line of reasoning renders the foundation of the experiments Bussemakers and de Haan have been carrying out [4, 5, 6, 7, 8, 9].

Bussemakers and de Haan use picture categorization as experimental task because this task resembles computer work to a certain extent: using modern computer interfaces usually boils down to categorizing different icons according to their possible significance. To represent the sounds used in such a computer interface, they present auditory distracters while participants carry out a picture categorization experiment. For auditory distracters Bussemakers and de Haan initially used earcons which are more abstract, constructed sounds, often very short musical fragments [10]. Later they also used auditory icons which are "caricatures of naturally occurring sounds" [11, p. 167]. In the picture categorization task, Bussemakers and de Haan requested participants to classify simple line drawings of animals and other objects according to the question "Is the picture you see that of an animal?" Bussemakers and de Haan presented the auditory distracters through headphones, without telling the participants that they could use the auditory information to their (possible) benefit.

According to the definition of redundant information, the auditory distracters were not meant to present new information nor were they meant to carry any information that could explicitly help the participants. Bussemakers and de Haan therefore use chords in major and minor mode because these carry a certain connotation by Western musical standards: Major chords are often said to evoke a positive and happy mood whereas minor chords evoke a negative and sad mood [12]. Participants hearing a major chord are inclined to respond positively to a question whereas participants hearing a minor chord are inclined to respond negatively more easily [13, 14, 15]. Bussemakers and de Haan therefore expected major chords to facilitate positive responses to the categorization task and inhibit negative responses. Minor chords were expected to facilitate negative responses and inhibit positive responses.

The first experiment [4] investigated the effects of earcons on picture categorization in a randomized design in an attempt to verify the above predictions. The expected facilitation and inhibition, however, was not found: Although response times to trials with sound were significantly slower than response times to trials without sound (on average 470 ms and 435 ms, resp.) the response times to trials with major chords versus trials with minor chords did not differ significantly. Bussemakers and de Haan concluded that participants were unable to relate the auditory distracters to the visual stimuli and reasoned that this could be caused by either the failure of the perceptual unification process or a perceptual overload of the auditory system due to the many different sounds employed in the experiment.

Bussemakers and de Haan [4] therefore used only one sound in their second experiment employing a Stimulus Onset Asynchrony (SOA), ranging from -500 ms to 0 ms, in their second experiment. The objectives of this experiment were two–fold: the reduction of the number of sounds to only one would reduce the chance of perceptual overload and the addition of the SOA allowed for the determination of the point of maximal interference of the auditory distracters on the picture categorization task. Finding this point of maximal interference automatically is important because at maximal interference, non–informational auditory distracters have their strongest effect on the picture categorization and participant thus have the best chance of using the information (if present) in the sounds to their advantage. If the findings of this experiment indicated that the maximal interference would arise at an SOA of 0 ms, this would enable Bussemakers and de Haan to conclude that perceptual overload had not occurred (because Experiment 1 [4] did not use an SOA, which, in practice, means an SOA of 0 ms) and that, perhaps, the perceptual unification had failed and had caused the absence of the expected facilitation and inhibition in the first experiment. The results of Experiment 2 [4] indeed indicated that the auditory distracters maximally interfered with the categorization task at an SOA of 0 ms.

A third experiment was set up to investigate if the failure to create a perceptual unity of the stimulus pair had caused the absence of the effects of inhibition and facilitation in Experiment 1 [4]. A blocked design was used that fixed the relation between the auditory distracters and the categories of pictures for the duration of a block. The relation between the earcons and picture categories concerned the responses indicated by the earcons and the correct responses to the pictures. Combining the response tendencies of the chords (a major chord and a positive answer) with the response requirements to the categorization task (pictures of animals should get a positive answer) created two different picture–sound sets: one set where pictures of animals were combined with a major chord (a positive response requirement and inclination) and, vice versa, pictures of other objects combined with minor chords (a negative response requirement and response inclination). Bussemakers and de Haan called this set the *congruent* condition, because the visual stimulus and auditory distracter have an equivalent connotation: both require or suggest a positive response. In the other set, response requirement and response inclination were mirrored: Pictures of animals were combined with minor chords and non–animal pictures were combined with the major chords. This set was called the *incongruent* condition. They also added two reference conditions. The baseline condition of the experiment consisted of pictures that were presented without sound. The reference condition for the congruent and incongruent condition was called the *neutral* condition and consisted of all pictures combined with the same sound (i.e., the pictures of animals had the same sound as the pictures of the other objects). Bussemakers and de Haan reasoned that such a blocked design would enable participants to deduce the relation, although subconscious, between the earcons and pictures more easily, and therefore now expected to find effects of inhibition and facilitation.

The acquired data clearly indicated that the participants used the contingency between the earcons and pictures, because the response times in the congruent condition and incongruent condition differed significantly. As in the first experiment [4], the conditions with sound were significantly slower than the condition without sound, but in this third experiment the incongruent condition was even slower than the congruent condition. The information in the earcon contradicting the required response to the picture, seemed to slow down the response times with an additional 20 ms, possibly because the human cognitive system had to choose the correct response from the two possible responses: the incorrect response indicated by the earcon and the correct response required to the visual stimulus. This may be refered to as an internal sanity check. The congruent and neutral condition only differed by approximately 3 ms, which may come as a surprise because of the different informational content of the auditory distracters in the neutral condition relative to the congruent condition. Where the cognitive system cannot extract useful information from the earcons in the neutral condition, the information the earcons carry in the congruent condition is potentially useful. Ordering by informational content, the neutral condition, in hindsight, might have been expected to have average response times in between those of the congruent and incongruent condition.

Lemmens [6] tested whether the musical connotation of the earcons (major and minor chords) might have caused the effects found in the third experiment [4] by replicating that experiment, with the exception that the group of participants now was separated in a group of musically skilled participants (with 6 or more years of experience playing a musical instrument) and a group of musically unskilled participants (without or with less than 6 years of experience). The findings show that the musical connotation of the earcons did not seem to matter: although the subgroup of musically skilled participants was on average around 30 ms faster in its responses to the categorization task than the group of musically unskilled participants, that difference did not approach the level of significance. These new findings provide additional evidence in favor of the earlier findings.

The apprehensive reader has probably already concluded that this task setup resembles a Stroop task [16], which is known for the strong interference occurring when subjects have to name the incongruent ink color of color words (e.g., the word "red" printed in green ink). Although the original Stroop task used stimuli that were presented in only one modality (vision), several multi–modal varieties of the Stroop task have been used, according to MacLeod [17]. These multi–modal Stroop experiments all show comparable interference on the primary task, of a component of the stimulus that cannot be ignored. In their experiments Bussemakers and de Haan assume that their interference is similar to that of the Stroop effect and that the obligatory processing of the auditory distracters —because the human ear cannot ignore stimuli— causes the effect of the earcons on the picture categorization. The different effects of the congruent and incongruent condition are being caused by the different relations the earcons and pictures have in each condition. This is evidence that indicates that the sounds do penetrate the cognitive level of analysis.

Bussemakers and de Haan reasoned that the co–occurring earcons apparently triggered a secondary task–set [18] because the obligatory processing of the earcons seemed to influence categorization in a top–down fashion. This secondary task–set initiates a response selection process as well. The cognitive system now has to select the correct response from this set of two possible responses, although this is only needed in the incongruent condition. This can be compared to the classic Stroop task in which participants have to suppress the automatically triggered task–set of word–reading and instead have to name the incongruent ink color of the words. The word–reading task–set causes many errors and an enormous slowing down of the naming task, resulting in Stroop interference. Although the interference effect of the Stroop task is much stronger and has been investigated more often, the Stroop task also has a facilitatory component. The facilitation is caused by the color words that are congruent with the ink names[1].

Relating their experiments to the Stroop task, Bussemakers and de Haan found two new opportunities for new experiments. They observed that the picture categorization task is a relatively simple task and that participants did not make many errors categorizing the pictures. The first opportunity for new experiments therefore was an investigation into the amount of errors in the picture categorization task with auditory distracters. The second series of new experiments involved the finding of a possible facilitatory component.

To investigate errors as a measure of performance and analysis, Lemmens, Bussemakers, and de Haan [8] designed a dual–task experiment in which participants had to carry out the usual picture categorization task as well as a mental addition task. The mental addition task was added to increase the amount of errors to enable statistically valid conclusions on the number of errors made. Response time analyses showed response time patterns comparable to those in the previous experiments [4, 5, 6]. Error analyses, however, revealed the remarkable pattern that the fewest errors were made in the congruent and incongruent condition, and the most in the silent condition. It seemed that participants were able to use the information in the earcons to their advantage, although the earcons themselves caused slower response times.

[1]Although two effects may be confounded as MacLeod [19] hypothesizes.

A further series of experiments on the effects of auditory distracters on picture categorization used auditory icons instead of earcons on the grounds that auditory icons have a more direct mapping to the pictorial stimuli [11] and thus are more likely to facilitate the categorization process. The first experiment in this series [7] employed a randomized design, as in the first experiment Bussemakers and de Haan carried out [4]. Instead of two earcons representing the two classes of pictures, Bussemakers and de Haan now used all the distinct sounds of all pictures of animals and musical instruments[2]. The trials in this experiment, therefore, comprised four different sets: one set, called "same", in which each picture was presented with its own corresponding sound. The second set consisted of all trials in which pictures of animals were combined with a sound of an animal and all musical instruments were combined with the sound of an instrument. This condition was called the "same category" set. These two sets can be said to be an equivalent of the congruent conditions in the earcon experiments[3]. Set 3 was the equivalent of the incongruent conditions: all pictures of animals were combined with a sound of a musical instrument and vice versa. This set was dubbed "other category". The final set consisted of the trials without sound and formed the baseline condition.

The results showed faster response times when the pictures were presented in conjunction with their own distinctive sounds relative to the baseline condition without sound. Comparing the stimuli with sound, Bussemakers and de Haan [7] did not find a statistical difference between the response times to "same" and "same category" trials, which they explained by the observation that the "same" trials are a subset of the "same category" trials. Combining these data and comparing them to the "other category" trials revealed a significant difference in response times (approx., 415 ms vs. 430 ms, resp.).

These results clearly present a response time pattern that is completely different from the response time patterns Bussemakers and de Haan found, using earcons as auditory distracters [4, 5, 6, 8]. Where earcons slow down the responses to the categorization task, auditory icons seem to facilitate the categorization task, resulting in faster responses (both observations are relative to the condition without sound). Because the main experimental task has hardly changed over the series of experiments Bussemakers and de Haan have been carrying out, only one possible conclusion presents itself: the differences in underlying structure and processing of the earcons and auditory icons causes the differences in the effects these two classes of sounds have on picture categorization. The more natural, symbolic relation [11] of the auditory icons and the pictorial stimuli perhaps enables an easier integration of these information resources, whereas the more abstract earcons require more processing to deduce their relation to the pictures.

The difference in the underlying structure of auditory icons and earcons, seems to reflect the different ways in which information is stored in human memory. Most information in our memory is either categorical or conceptual in nature. Whereas categorical knowledge, that is perceptual in nature, enables humans to generalize by comparison to objects or events they encountered before, conceptual knowledge allows for generalization by the use of rules. For example, when deciding what to do when a barking

[2]Musical instruments specifically, instead of other objects, were used to simplify the creation of appropriate auditory icons.

[3]Because of this equivalency, the terms congruent and "same (category)", and incongruent and "other category", in the context of auditory icons, will be used interchangeably in the remainder of this paper.

dog is heard, humans use their categorical knowledge more than their conceptual knowledge. On the other hand, when deciding if a number is odd or even usually the rule 'divisible by two' is used, which requires almost no categorical knowledge.

Bussemakers and de Haan linked this theory of perceptual versus conceptual knowledge, also known as the what/where distinction or direct perception versus representation [20], to their experiments by assuming that two different categorization processes exist as well. One process categorizes objects according to rules and another process categorizes by comparison to examples or prototypes. Their assumption is supported by research carried out by Warrington and Smith [21, 22]. Warrington, investigating categorization in patients with left– or right–hemisphere lesions, found an impairment in patients with right–hemisphere lesions for perceptual categorization tasks, whereas patients with left–hemisphere lesions were impaired on tasks with a greater semantic (and thus conceptual) component. This led Warrington to pose two different categorization processes: one rule–based process and a process categorizing using examples, dubbed the theory of two–stage categorization. Smith found comparable differences in normal subjects as well, both for artificial as well as natural objects and also claimed two categorization processes.

By assuming that two categorization processes exist, the inhibiting effect of the earcons on response times to the categorization task, can be explained by assuming that earcons have to be processed by the rule–based process, because most participants in the Bussemakers and de Haan studies have not been trained musically to such an extent that recognition of chords occurs by comparison to previously encountered examples. This forces the cognitive system, that unifies the stimuli in the auditory and visual modality, to wait for the finishing of the auditory categorization process as the categorization of the simple pictures has already finished because of the faster example–based process [22]. Auditory icons on the other hand, are encoded by the example–based process as well and can therefore be expected to have different effects on the picture categorization.

Now two new experiments will be presented and, in the general discussion, the results will be summed up and a general conclusion based on these new, and the older, results will be presented. The first experiment that will be discussed, uses auditory icons instead of earcons to replicate the experiment the authors presented at ICAD2000 [8]. Experiment 2 is an experiment to investigate the effects of auditory icons and earcons on picture categorization when these sounds are used together in one design.

2. EXPERIMENT 1: AUDITORY ICONS IN A DUAL–TASK EXPERIMENT

This experiment was primarily set up to acquire data that allowed for error analyses in the auditory icon context. Furthermore, because Bussemakers and de Haan did not use a strictly blocked design in the first study [7] with auditory icons as in [4, 5, 6, 8], they needed an experimental setup with auditory icons that more strictly resembled that when using the earcons.

These requirements resulted in the construction of the present experiment. To meet the first requirement, again a dual–task design was constructed with a secondary mental addition task. By using auditory icons by category (e.g., the sound of a dog barking for all pictures of animals) instead of per picture, a blocked design with congruent and incongruent conditions was created. In addition to the sound of the barking dog, the sound of a piano playing was used to represent the category of other objects (in this experiment musical instruments). This design also met the second requirement.

The earlier experiments [7] with auditory icons in a single–task situation showed that having the redundant auditory information leads to shorter reaction times when compared to a situation with no auditory information. It seems that there is an alerting effect of the auditory stimulus to the visual stimulus. Furthermore if the information is of the same category as the visual information, subjects benefit from having the extra information as can be seen in a further facilitation effect.

In the experiment presented here however, there is a dual–task setting, that, similar to earlier dual–task experiments with earcons [8], could lead to a general inhibition of the reaction times. Because subjects want to make as few errors as possible and part of the cognitive capacity is used by the addition–task, overall the reaction times could be longer. On the other hand it is also possible that the effect is different for conditions where the category of the sound is congruent with the category of the pictures than for conditions where the category of the sound is incongruent with the category of the pictures.

Finally, and this is one of the reasons for conducting this experiment, it is important to evaluate the error–rates. What if the reaction times in the dual–task situation are shorter in the conditions with sound than in the conditions without sound, but subjects make more errors? In certain environments where a low number of errors is critical, it can be important to also know the effect of the auditory icons on error rates.

2.1. Method

2.1.1. Participants

Twenty students of the University of Nijmegen participated in the experiment. They were paid or received course credits for their time.

2.1.2. Materials

The experiment was carried out on a Macintosh Quadra 840 AV. The visual stimuli were presented on a 14–inch screen that was raised to eye–level. The auditory stimuli were presented via a pair of Monacor BH–004 headphones. The responses were registered through a button–box. The visual stimuli consisted of 7 line drawings of animals and 7 line drawings of musical instruments (see Table 1) The pictures were modified to include a one–place number in a position close to the center of the drawing (see Figure 1).

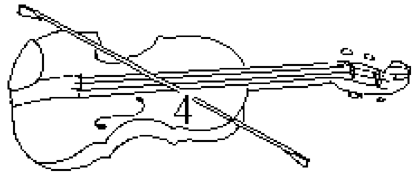

Figure 1: *Example visual stimulus for Experiment 1.*

Three sounds were used as auditory stimuli. The sound of a dog barking represented the animal category; for the category of musical instruments the sound of a piano playing was used. The sound of dripping water was used as a neutral stimulus. None of the sounds were represented in the pictures. The duration of the sounds was normalized to 1227 ms.

Table 1: *All visual stimuli in Experiment 1.*

Animal	**Musical instrument**
Duck	Flute
Donkey	Guitar
Cat	Harp
Chicken	Organ
Horse	Drum
Lion	Trumpet
Cow	Violin

2.1.3. Design

To allow the subject to take full advantage of the relationship between the picture and the sound, 6 conditions were defined (see Table 2). Furthermore because of the similar setup, the results could be compared to earlier studies with earcons (see also [4, 5, 6, 8]). In the congruent condition the pictures of animals were presented with the sound of a dog barking. With the pictures of musical instruments the sound of someone playing the piano was presented. Both the sound and the picture suggest the same response, because they are of the same category. In the incongruent condition the opposite was the case: With pictures of animals the subjects heard the piano sound and with pictures of musical instruments the sound of a dog barking. The visual and the auditory information suggest a different response. In the silent condition the pictures were presented with no additional auditory stimulus.

Table 2: *All experimental conditions in Experiment 1. In the congruent condition the pictures of animals were accompanied by the sound of a dog barking and all other pictures were accompanied by the sound of a piano playing.*

	Picture category	
Condition	Animal	Musical Instrument
Congruent	Dog	Piano
Incongruent	Piano	Dog
Neutral	Piano	Piano
	Dog	Dog
	Water	Water
Silent	–	–

There were three neutral conditions. In previous studies the neutral condition was defined as the condition where for both categories of visual stimuli the same auditory stimulus is presented. However in the case of auditory icons, the perceptual, categorical nature of the sound provides information not only on a condition–level, but also on a trial–level. It is possible that in the neutral condition with the sound of the dog, the pictures of animals benefit more from the additional auditory information than the pictures of the musical instruments. Therefore a third neutral condition was included with a sound that is not related to any of the categories.

To ensure that every condition followed every other condition an equal number of times, a digram–balanced Latin Square (e.g., Wagenaar [23]) was used to control for any order effects. Furthermore the position of the buttons was varied across subjects, to counterbalance for any effects of prefered hand.

2.1.4. Procedure

Participants carried out two tasks of which the order was controlled. The first task was a visual categorization task with additional auditory distracters. In the second task, participants were also instructed to accumulate, per category, the numbers presented in the pictures. Each trial started with the presentation of a fixation cross and sound, for 500 ms. Then the sound and picture were presented. The picture was displayed for 300 ms. Subjects could indicate their response to the question whether or not they saw a picture of an animal by pressing a button labeled 'yes' or a button labeled 'no'. A new trial was started after 2500 ms.

2.2. Results and Discussion

In the statistical analyses error–responses and no–responses were excluded. The mean reaction times for both tasks are shown in Figure 2.

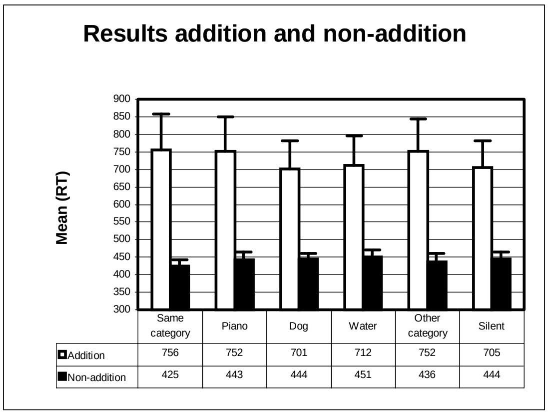

Figure 2: *Average response times (ms) per condition in Experiment 1.*

A repeated measurements analysis indicated that the conditions with the extra addition task were significantly slower than the conditions without the extra addition task ($F(1, 19) = 15.624$, $p < .001$).

Within the non–addition task no significant differences between all conditions with sound (including the neutral conditions) and the silent condition were found ($F(1, 19) = 1.385$, $p > .1$). However when comparing the results in the congruent condition together with the incongruent condition to the results in the silent condition, these two were significantly faster ($F(1, 19) = 6.711$, $p < .05$). It seemed that when the sounds were different for each category, the reaction times were faster than when there was no sound present.

The mean reaction times in the neutral conditions did not differ significantly from either the incongruent condition or the silent condition ($F(1, 19) = 1.298$, $p > .1$ and $F(1, 19) = 0.195$, $p > .5$). When comparing the neutral conditions to the congruent condition, the congruent condition was significantly faster ($F(1, 19) = 9.180$, $p < .01$).

Within the dual–task situation, the mean reaction times on the conditions with sound were significantly slower than the silent condition ($F(1, 19) = 4.792$, $p < .05$). Comparison between the mean reaction times on conditions in the dual task situation showed trends, but no significant differences.

Analyzing the errors showed that in 3.1% of the trials an error was made. Furthermore 53% of those errors were made in the addition–task and 47% of the errors were made in the non–addition task. In Table 3 the mean proportion of errors per task and per condition is displayed. A repeated measurements analysis did not show a significant difference in mean proportion of errors between tasks ($F(1, 19) = 0.479$, $p > .4$). Within each task, none of the differences in proportion of errors between conditions were significant.

Table 3: *The average number of errors per task per condition. The amount of errors is averaged over the total number of trials. The number in braces is the standard error.*

	Average number of errors	
Condition	**Addition**	**Non–addition**
Congruent	0.021 (0.009)	0.021 (0.007)
Piano	0.028 (0.009)	0.028 (0.009)
Dog	0.039 (0.012)	0.032 (0.010)
Water	0.025 (0.011)	0.018 (0.007)
Incongruent	0.025 (0.009)	0.018 (0.007)
Silent	0.025 (0.005)	0.025 (0.004)

Earlier studies [7] show that having auditory icons in a visual categorization task can lead to faster response times, if the sounds are different for each category. Especially when the auditory information is of the same category as the visual information, subjects respond fastest.

The results from the non–addition task in this study confirm the results from earlier studies, that have shown that auditory icons in a visual categorization task can lead to faster response times, if the sounds are different for each category. Especially when the auditory information is of the same category as the visual information, subjects respond fastest. It seems that having categorical auditory information, even if it is not the same as the visual information, leads to faster responses, because one stimulus seems to alert to the other. Between the neutral conditions and the silent condition there is no significant difference. Having the same auditory information with every picture within a block, does not seem to influence the response times, regardless of the category of the picture. Whether this information is related to one of the categories, like, for instance, in the case of the dog sound or the piano sound, or an entirely different sound like the water, does not seem to matter. It seems that the auditory information needs to be different per category to assist in the response to the visual information.

In the addition task there is no significant difference between the conditions with sound, but there is a significant difference between the conditions with sound and the silent condition. Having extra auditory information while having to add numbers seems to slow the response down, but it does not matter what kind of information is in the sound. Just the fact that the auditory information needs to be processed and that this interferes with the mental addition can possibly explain the findings. Subjects need to first understand what the sound is before they can disregard it in the context of the task.

The error data shows that there is no difference in mean proportion of errors between conditions. It seems that the different types of additional auditory information do not lead to differences in error–rates.

From these findings it can be concluded that having auditory icons in a visual categorization task leads to shorter reaction times when compared to a situation where there is no sound. However if there is a secondary cognitive task, the effect changes and reaction times are slower in the conditions with sound compared to the silent condition. It seems that in a more complex situation, having additional information in another modality needs to be processed, which leads to a decrease of the reaction times. Comparing these results to earlier findings on earcons, there is a clear difference. The reaction time data shows that contrary to auditory icons, earcons, both in a single–task and a dual–task setting, lead to longer reaction times.

These findings are restricted to categories of animals, musical instruments, and other objects. It seems interesting to find out if the results would be similar if other categories were used. Also, the categories that are used here are concrete and perceptual. It is possible that the results would be very different for abstract categories, for instance with a category like emotions.

3. EXPERIMENT 2: INTERMIXED AUDITORY ICONS AND EARCONS

Because no experiments combining earcons and auditory icons in one experimental setup have been carried out, Experiment 2 was set up to investigate whether this combining would result in different effects than the known effects of earcons and auditory icons. The experimental setup used is a combination of Experiment 1 [this paper] and the experiment described in [8], with several small modifications. The addition task from Experiment 1 and [8] was not replicated and the sets of visual stimuli from the previous experiments were mixed and suitably adapted for use in a single design.

The changes to the experimental stimuli resulted in two sound categories (auditory icons and earcons) and three sets of pictures (animals, musical instruments, and other objects). The experimental design consisted of five conditions (four experimental condition and one baseline condition) in each of which the relation between a sound category and a picture category was fixed.

Because the previous effects of earcons and auditory icons on picture categorization seemed reasonably consistent, Lemmens, Bussemakers, and de Haan expected to find these effects in this experiments as well. Although the mixing of the earcons and auditory icons might have resulted in less pronounced effects within conditions with equivalent sounds, auditory icons were still expected to facilitate response times whilst earcons were expected to slow down response times.

This experimental setup, however, has one downside: the class of non–target pictures cannot be the same across the sound categories. Because Bussemakers and de Haan used the sound of a piano playing to represent the non–target pictures in the auditory icon experiments, the experimental setup in those experiments required the use of pictures of musical instruments instead of other objects[4]. This may cause an unwanted effect because the category of musical instruments is a subclass of the category of other objects and because the response times to musical instruments will be, a priori, faster than the responses to the pictures of other objects (due to the use of auditory icons versus earcons), the non–target pictures can, and may, not be compared directly. However, Bussemakers and de Haan did not expect this unwanted effect, because of the blocked experimental design: participants will probably not notice the different non–target categories between the blocks.

3.1. Method

3.1.1. Participants

Twenty students (11 female, 9 male; mean age = 22.8 years) of Cognitive Science or Psychology at the University of Nijmegen participated in the experiment. They were paid 5 guilders or received course credit for their participation. None of these students have taken part in the previous Bussemakers and de Haan studies [4, 6, 7, 8, 9].

3.1.2. Materials

The materials from the present study have been used in two previous studies [8, 9] and included visual (pictorial) stimuli and auditory distracters. The original set of target pictures consisted of 11 black and white line drawings of animals (e.g., a lion, a cat, and a butterfly); the original set of non–target pictures comprised and equal number of pictures of musical instruments and other objects (e.g., a trumpet, a violin, and a candle). The original set of auditory distracters consisted of a C–major chord and a C–minor chord (triads with C5 as tonic), the sound of a barking dog, the sound of a piano playing, and the sound of splashing water. These sounds had different durations.

To create the stimulus sets for the present study, the original sets were changed in several ways. A number of pictures that caused problems in earlier studies were removed. The sound of splashing water was removed from the set of auditory distracters; the remaining auditory distracters were normalized in their duration to that of the sound of the barking dog.

The final set of experimental stimuli incorporated 21 pictures (7 pictures of animals, 7 pictures of musical instruments, and 7 pictures of other objects), and 5 sounds (an alert sound, a C–major chord, a C–minor chord, the sound of a dog barking, and of a piano playing). The pictures all had the same size; all sounds (with the exception of the alert sound, which was a simple beep) had durations of approximately 1200 ms.

3.1.3. Design

The stimuli were presented in a blocked within–subjects design. The blocks were characterized by the relation between the visual stimuli and the auditory distracters. Blocks containing the C–major and C–minor chord were characterized as "abstract". Blocks containing the piano and the barking sound were called "concrete". Within these classifications, blocks were further categorized as congruent and incongruent. In congruent blocks, the visual stimuli and auditory distracters had equivalent connotation, for example, a picture of a cat combined with the barking sound or the C–major chord. In incongruent blocks the visual stimuli had an opposite connotation compared to the auditory distracters, for example, the picture of a cat combined with the sound of a piano or a C–minor chord. This design resulted in four different blocks which were presented, properly counterbalanced, in a digram–balanced Latin square design.

Within each block all visual stimuli were presented four times: twice with sound and twice without. Given a set of 14 pictures per block (7 pictures of animals and 7 of musical instruments or other objects), this resulted in 56 trials per block, which in turn implied 224 stimuli overall (practice set not included).

[4]which have never included musical instruments.

3.1.4. Procedure

The stimuli were presented on a Macintosh Quadra 840AV computer, equiped with a 256–color screen, with an effective diameter of 32 cm, and linked to a buttonbox and a pair of headphones. The visual stimuli were presented in a centered white square of 18 cm wide and 12 cm high. The auditory distracters were presented to both ears, but without any stereophonic effects.

Each trial started with the presentation of an alert sound and a fixation cross in the center of the display, for 500 ms. After a 500 ms pause, the visual and auditory stimulus pair was presented. The visual stimulus was presented for 300 ms and had the same onset as the auditory distracter. The maximal response time was set at 2500 ms. The intertrial interval lasted for 1500 ms.

Participants were instructed to press one button as fast and accurately as they could if the picture presented on the computer screen was that of animal, and another button if the picture was not that of an animal. Participants first trained on a set of 28 stimuli from the experiment set. The practice set covered a randomized subset of a complete set of blocks. Between this set and the actual experiment a pause of at most a minute was inserted. All experimental blocks were separated by a voluntary break of at most a minute as well. An experimental session lasted for approximately 30 minutes.

3.2. Results and Discussion

Before any statistical analyses were carried out, the incorrect as well as the null response were removed. These comprised 2.05% of the total number of trials. The average response times are presented in Figure 3.

A repeated measurements analysis showed that both the response times in the concrete conditions and in the abstract conditions were significantly different from those in the silent condition. Concrete conditions were significantly faster ($F(1, 19) = 10.473$, $p < .05$) and abstract conditions were significantly slower ($F(1, 19) = 8.507$, $p < .05$).

Within the concrete conditions, however, the differences between the congruent and incongruent condition did not reach the level of significance ($F(1, 19) = 2.272$, $p > .1$). This applied for the abstract conditions as well ($F < 1$).

These results provide additional evidence in favor of the findings of [6, 7, 8] that earcons have an inhibitory effect on response

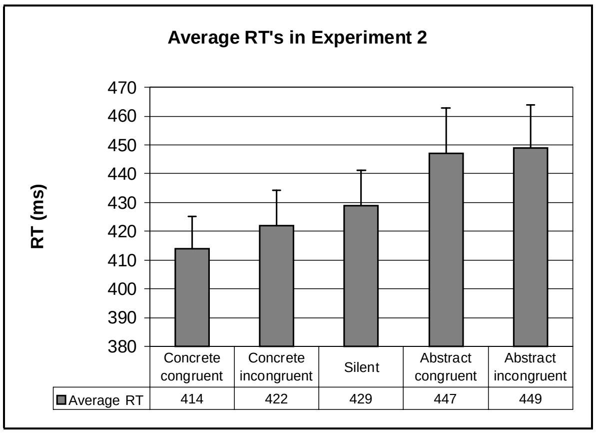

Figure 3: *Average response times (ms) for all conditions in Experiment 2.*

times to a visual categorization task, whereas auditory icons have a facilitating effect. The present results furthermore indicate that auditory icons and earcons can be used intermixed without a great chance that new effects instead of the usual facilitation and inhibition may occur.

The experimental design, however, has not enabled us to check whether a subset effect of the musical instruments being a subset of the other pictures occurred. Correct statistical analyses to check for this effect require a different experimental setup, which would be a suggestion for further research.

4. GENERAL DISCUSSION

The data from Experiment 1 and Experiment 2 again consistently show that auditory icons facilitate response times to a visual categorization task whereas earcons have an inhibitory effect. The effects are of course limited to categorization experiments employing pictures of animals in combination with pictures of musical instruments and other object categories. Although this limitation certainly exists, Lemmens, Bussemakers, and de Haan strongly feel that the consistency of the results suggests that other experimental paradigms and other sets of sounds and pictures are likely to show the same pattern of responses, provided that equivalent congruent and incongruent relations exist.

Experiment 1 shows that in a dual task setting a pattern of response time latencies is found that is comparable to [7] and [8], although the data from Experiment 1 show that, in contrast to [8], the auditory icons have an inhibiting effect in the addition task. This finding, however, can possibly be explained by assuming that the addition task has an interfering effect on the processing of the auditory icons. The error analyses show a trend comparable to that found in [8]. The differences between the amount of errors in the individual conditions, however, do not approach the level of significance. It may be that this is caused by the still very low amount of errors in Experiment 1, despite the addition task. Although these studies are different both in construction and intricacies of results, these findings do support each other and add to the robustness of the effects of auditory icons.

Experiment 2 shows that auditory icons and earcons can be readily intermixed in an experiment without compromising the strength of their individual effects on the primary categorization too much. Although the differences between the congruent and incongruent conditions in both the concrete and abstract conditions are statistically not significant, the findings of Experiment 2 corroborate the findings of all previous studies [4, 5, 6, 7, 8, 9].

The effect the added earcons have on the response times to the categorization task is always one of slowing down those response times. The auditory icons usually have a facilitating effect on response times to the categorization task. These effects can be further differentiated when taking the congruency of pictorial stimulus and auditory distracter into account. In general, conditions with a congruent relation between picture and sound are faster than conditions in which an incongruent relation is employed. The congruent condition, however, never differs more than a few milliseconds from the neutral condition, suggesting that the difference between the incongruent and congruent condition is likely to be caused by a sanity check in the incongruent condition, to verify the correct response out of a set of two possibilities. The different effects earcons and auditory icons have on visual categorization must be caused by differences in their structure and processing, because the primary task participants carry out does not change greatly across experiments.

Earcons, in general, as well as in these studies, are often more abstract, constructed sounds [10]. Although western music is filled with major and minor chords, and most participants have experience in listening to them, most participants are not trained to automatically recognize them. They experience the chords as they are, not for what they are and therefore they probably need rules in order to name the chords; in the domain of music psychology these rules usually are formulated in terms of the findings of Crowder [13, 14, 15]. On the other hand, auditory icons are caricatures of naturally occurring sounds [11], which most people recognize directly by comparing them to the prototypical examples which they have encountered before.

Processing earcons according to rules and auditory icons by prototypical example is reminiscent of the alternative categorization process posited by Warrington and Smith [21, 22], comprising two subprocesses. The perceptual process categorizes objects by comparison to a stored (prototypical) examples whilst the conceptual process encodes objects by rule application. Because rule based encoding involves extensive use of working memory and selective attention, it is a relatively slow process [22]. Smith tested both naturally occurring and artificial sounds and found that both sound categories adhere to the theory of two–stage categorization.

Assuming that the processing of the visual stimuli confirms to the two–stage categorization theory [21] as well, the differences between the effects of the auditory icons and earcons might be explained by relating the encoding process of the auditory component of a stimulus pair and the visual component of that pair. Encoding the line drawings that were used in the experiments does not explicitly require the use of rules to determine whether a certain picture does or does not belong to the class of animals. According to the two–stage categorization process, the line drawings therefore are encoded by example–based categorization. Because auditory icons are encoded by the same example–based categorization subprocess, they may be expected to facilitate (or at least not hinder) the picture categorization, because a faster integration of information resources in different modalities is possible. Earcons, on the other hand, are encoded by the relatively slow rule–based categorization subprocess. The cognitive system has to wait for both categorization processes to finish (one process for the auditory distracter and one for the visual stimulus) and this waiting results in slower response times.

This distinction between rule–based categorization of earcons and example–based categorization of pictorial stimuli and auditory icons is of course a specification of a more general theory of perceptual and conceptual knowledge. This theory claims that human memory retains information either as categorical (perceptual) or conceptual knowledge. Categorical knowledge, which is perceptual in nature, enables us to generalize about what we have learned about an object or an event to other similar objects or events and the generalization allows us to determine the appropriate means of action, for example, determining how it can be that you see a dog and a cat chasing each other, while at the same time hearing quacking ducks. Categorical knowledge is limited, however, because it depends on the perceptual similarity between the object and the previously encountered objects. Conceptual knowledge, on the other hand, can help in setting the boundaries of what belongs to a category and what does not, by the use of principles and rules. Concepts are based on deeper and more abstract properties of objects. For instance, determining if a number is odd or even requires the verification if the number can be divided by two.

Summarizing the findings so far it can be seen that auditory icons and earcons have different effects on the same task. The experiments employing earcons as auditory distracters show that the earcons slow down the response times to the categorization task whereas experiments employing auditory icons show that these facilitate response times. Although the strength and significance of these effects differs per experimental setup, the general effect of inhibition or facilitation can be verified in each of the experiments. This of course limits the conclusions to just this task, but the blocked design enables participants in the experiments to quickly use the relation between sound and picture and to use this relation without being consciously aware that the relation exists. Therefore it is expected that new experiments using a different task within a comparable blocked design will show comparable results, both on the level of response time latencies as well as on the level of amount of errors.

Having discussed these new results and experiments and in relation to the older experiments, it is necessary to suggest some design considerations when implementing sound in a user interface. First, modesty and a planned approach towards the use of sound in interfaces are very important: Although sound may seem to lead to performance enhancements [1], sound may also annoy the users of the interface and it is often unclear if the information the sound presents or the presence of the sound itself causes the performance enhancements [2]. Furthermore, the importance of the relation between the sound and the (visual) event it accompanies needs to be emphasized: in time–critical environments employing earcons in incongruent relations must be prevented, because the incongruent conditions in the earcon experiments have consistently shown to be slower than the congruent, neutral, and silent conditions. Finally, it must be pointed out that the consistent use of the relation between auditory icons and earcons and visual events may lead to a decrease in faulty decisions; this conclusion, however, can only be presented tentatively, because strong statistical evidence lacks.

Although much work can still be done in just this area (priming techniques, fMRI, EEG, etc.), the prospects for new experiments are drifting away from the techniques described in this paper. The new experiments will probably involve spatial Stroop tasks and Simon tasks, constructed from the perspective of executive control and action plans.

5. REFERENCES

[1] M. Rauterberg, "About the importance of auditory alarms during the operation of a plant simulator," *Interacting with Computers*, vol. 10, pp. 31–44, 1998.

[2] J. Edworthy, "Does sound help us to work better with machines? A commentary on Rauterberg's paper 'About the importance of auditory alarms during the operation of a plant simulator'," *Interacting with Computers*, vol. 10, pp. 401–409, 1998.

[3] S. A. Brewster, "Sound in the Interface to a Mobile Computer," in *Human–Computer Interaction: Ergonomics and User Interfaces, Volume 1 of the Proceedings of the 8th International Conference on Human–Computer Interaction*, H.–J. Bullinger and J. Ziegler, Eds., Londen, 1999, vol. 2, pp. 43–47, Lawrence Erlbaum Associates.

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

[5] M. P. Bussemakers and A. de Haan, "Using earcons and icons in categorisation tasks to improve multimedia interfaces," in *Proceedings of ICAD'98*, Glasgow, UK, 1998, The British Computer Society.

[6] 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.

[7] M. P. Bussemakers and A. de Haan, "When it sounds like a duck and it looks like a dog… auditory icons vs. earcons in multimedia environments," in *Proceedings of the International Conference on Auditory Display*, P. R. Cook, Ed. 2000, pp. 184–189, International Community for Auditory Display.

[8] 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.

[9] M. P. Bussemakers and A. de Haan, "Auditory icons and earcons: Categorical and conceptual multimodal interaction," Manuscript submitted for publication.

[10] M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg, "Earcons and icons: their structure and common design principles," *Human–Computer Interaction*, vol. 4, pp. 11–44, 1989.

[11] W. W. Gaver, "Auditory icons, using sound in computer interfaces," *Human–Computer Interaction*, vol. 2, pp. 167–177, 1986.

[12] K. Hevner, "The mood effects of the major and minor modes in music," in *Proceedings of the Midwestern Psychological Association*, 1933, p. 584.

[13] R. G. Crowder, "Perception of the major/minor distinction: I. Historical and theoretical foundations," *Psychomusicology*, vol. 4, no. 1–2, pp. 3–24, 1984.

[14] R. G. Crowder, "Perception of the major/minor distinction: II. Experimental investigations," *Psychomusicology*, vol. 5, no. 1–2, pp. 3–24, 1985.

[15] R. G. Crowder, "Perception of the major/minor distinction: III. Hedonic, musical and affective discriminations," *Bulletin of the Psychonomic Society*, vol. 23, no. 2, pp. 314–316, 1985.

[16] J. R. Stroop, "Studies of interference in serial verbal reactions," *Journal of Experimental Psychology*, vol. 18, pp. 643–662, 1935.

[17] C. M. MacLeod, "Half a century of research on the Stroop effect: An integrative review," *Psychological Bulletin*, vol. 109, no. 2, pp. 163–203, 1991.

[18] R. D. Rogers and S. Monsell, "Costs of a predictable switch between simple cognitive tasks," *Journal of Experimental Psychology: General*, vol. 124, no. 2, pp. 207–231, 1995.

[19] C. M. MacLeod and P. A. MacDonald, "Interdimensional interference in the stroop effect: uncovering the cognitive and neural anatomy of attention," *Trends in Cognitive Sciences*, vol. 4, no. 10, pp. 383–391, 2000.

[20] U. Neisser, "Multiple systems: A new approach to cognitive theory," *European Journal of Cognitive Psychology*, vol. 9, no. 3, pp. 225–241, 1994.

[21] E. K. Warrington and A. M. Taylor, "Two categorical stages of object recognition," *Perception*, vol. 7, pp. 695–705, 1978.

[22] E. E. Smith, A. L. Patalano, and J. Jonides, "Alternative strategies of categorization," *Cognition*, vol. 65, pp. 167–196, 1998.

[23] W. A. Wagenaar, "Note on the construction of digram–balanced Latin squares," *Psychological Bulletin*, vol. 12, no. 6, pp. 384–386, 1969.