The Effect of Earcons on Reaction Times and Error-Rates in a Dual-Task vs. a Single-Task Experiment

Paul M.C. Lemmens, Myra P. Bussemakers, Abraham de Haan
Nijmegen Institute for Cognition and Information
University of Nijmegen
Montessorilaan 3,
6525 HR Nijmegen, The Netherlands
+31 24 3612648
bussemakers@nici.kun.nl

ABSTRACT
An experiment with two picture categorization tasks with auditory distracters containing redundant information was carried out to investigate the effects distracters, in this case earcons, have on categorization. In the first task participants had to carry out an extra mental addition task. The secondary task consisted of just the categorization. The dual-task situation was expected to lead to longer reaction times and more errors than the single-task situation, possibly providing new insights when analyzed. The results confirmed previous experimental results and indicated that significantly fewer errors were made in conditions in which the earcons contained relevant redundant information, compared to conditions with no relevant redundant information.

Keywords
Earcons, auditory distracter, multimodal interfaces, categorization, dual-task performance

INTRODUCTION
One of the areas in human-computer interaction and interface design that is more and more under investigation is multimodal interaction. Researchers as well as designers want to learn what the consequences are of communicating information to the user through the visual, auditory or even the haptic modality. Learning about the consequences of communicating (different) information to different modalities first requires an understanding of how users of a multimodal interface relate the information presented to those modalities. Are users even aware that information presented in one modality has a relationship with information presented in another modality? To investigate this question, usually the notion of contingencies is used.

A contingency describes a more or less invariable feature in separate flows of information, which allows the human cognitive system to relate these flows of information to acquire new information. For instance, an audience enjoying the talents of a ventroquilist is led to believe the puppet of the ventroquilist is talking like it has a mind of its own, while actually the puppetplayer is making the speech-like sounds and the puppet only mimics the lipmovements of speech. Although both information sources are presented to different modalities, most people in the audience will "hear" the puppet talking. The audience apparently relates the information presented in the auditory modality to the information presented in the visual modality, because the sources of that information are in close proximity and the flow of information runs synchronously; the contingencies for proximity and synchrony are used.

Contingencies therefore present opportunities to relate different sources of information to each other and use that relationship to unite the information into an information unit that can be evaluated by the cognitive system as being singular. The effects or consequences this singular unit of information has on performance in a multimodal environment depends on the actual informational content of the separate information sources, because the information presented in the additional modality can complement the information in the main modality (i.e., provide extra information not presented already). However, the additional information can also have a redundant nature (i.e., presenting information already present in the main modality). It seems likely that complementary information has different consequences for information processing than redundant information has.

Extra information can be expected to have a positive effect on performance on tasks within a multimodal interface because, for example, the new additional information provides a means to decrease errors. Earlier research has shown that if auditory information is used to supplement visual information, for instance because the eyes are otherwise occupied or if it is difficult to perform the task with only the visual information, sounds can lead to faster response times and fewer errors [2]. This is a relatively self-evident effect. It is debatable, however, whether redundant information produces comparable effects.

In this project therefore, the effect of redundant auditory information on a primary visual categorization task is investigated. This means that the information that is presented through audio is also available visually and users are able to perform the task, in principle, equally well without listening to the sound. In the experiment conducted in this project, users have to categorize line-drawings of animals and non-animals (e.g., a chair) by indicating whether the picture they see is that of an animal, "yes" or "no". At the same moment they see the picture, an abstract musical chord in major or minor mode (an earcon, [1]) is played. Furthermore users are not instructed to "use" the information in the sound when working on the task. However, they are told that the sounds they will hear can aid them in the categorization task.

Because the major and minor modes in western music have a fixed connotation (a piece in major mode is usually said to evoke a "happy mood" while a piece in minor mode is said to be "sad"), these connotations are expected to be carried over to the earcons. The earcon in major mode is therefore expected to evoke a positive mood in participants while the earcon in minor mode is expected to put participants in a sad mood. Because participants in a positive mood are known to respond quicker to questions requiring a positive answer, participants hearing a major earcon are expected to respond quicker to stimuli requiring a "yes"-response than to stimuli requiring a negative response [6]. Earcons in major mode are therefore expected to facilitate affirmative responses and inhibit disaffirmative responses, while earcons in minor mode are expected to facilitate disaffirmative responses and inhibit affirmative responses. The additional information earcons carry is then comprised of the mood the earcons evoke.

Experimental verification of these expectancies showed that only in experiments in which the relation between the earcon and the visual stimulus was fixed for a certain period of time, an effect could be shown [3]. However, the facilitation, or inhibition, as expected in its purest form, could not be shown. For instance, the effect of facilitation was not found, because in congruent conditions (in which the facilitating, major mode earcon was coupled to visual stimuli requiring a positive response) RT's did not differ from conditions in which the earcons carried neutral, additional information. The congruent condition however was expected to show decreased response time latencies, because of the facilitation [3,4,5].

The earlier results have shown though, that having additional information in the auditory modality can slow down the performance if the sound does not coincide with the visual information. For instance, when pictures have to be categorized as animal or non-animal, a minor chord presented together with a picture of an animal leads to slower responses compared to a congruent stimulus pair (e.g., major earcon and an animal picture, [4,5]). However, only looking at the speed of the responses does not give us a full picture of what happens in these multimodal situations. Errors also are a source of information on what processes take place on a cognitive level. Because in the previous categorization experiments subjects make very few errors, a second task was added to evoke more errors. The results of these tasks are presented in this paper.

The assumption that a secondary task can increase the cognitive load for participants is based on the theory that participants have a shared pool of mental resources available [7,8]. Carrying out two tasks simultaneously requires a larger amount of resources than the sum of the resources necessary for carrying out each task individually, which, among other things, can result in more errors in the performance. It is likely therefore, that participants, in an effort to maintain an acceptable degree of performance both in errors and response time latencies, try to carry out the tasks sequentially. This requires separate buffers for both the visual stimuli and the auditory distracters. In the model of Logie [7] the buffer for the visual stimuli is called the visio-spatial buffer while the auditory buffer is called the phonological loop. Using a temporary buffering strategy for carrying out both tasks in a dual task situation sequentially is usually reflected by an increase in response time latencies only and the absence of an increase in errors. Because participants try to minimize the number of errors they produce, they take a longer time to complete the two tasks, compared to a single task situation, and the prolonged response time latencies indeed result in comparable number of errors in the single-task and dual-task situation [8]. However, the overall number of errors will increase and these can be analyzed as well, next to the usual reaction time analyses, possibly providing new insights for the theoretical framework in which this experiment can be placed.

The theoretical framework surrounding this experiment is that of two-stage categorization. This theory states that the categorization process, usually viewed as a singular process, actually consists of two stages or processes defined by the way each sub-process categorizes items. One process categorizes by rules while the other categorizes by example. Rule based categorization usually is a relatively slow process because it involves extensive use of selective attention and memory [8]. Exemplar similarity based categorization, on the other hand, is a relatively fast process because it only requires direct comparison with a stored (prototypical) example of a category [9]. Because the musical chords, used in the experiment described here, are relatively abstract sounds for most participants, it is likely most of them need to use rules to process the earcons semantically. They should therefore be slower in conditions with sound compared to conditions without sound. More experimental results providing additional evidence in favor of two-stage categorization, can be found in another article by Bussemakers *et al.* elsewhere in these proceedings.

The musical connotation of the auditory distracters is also a reason to consider musical skill of participants, because musically trained participants might process the musical chords used in the experiment differently than musically untrained participants. Differentiation on musical skill could therefore show qualitative differences between participant groups. However, because the effects the auditory distracters have on the categorization task are largely subconscious, it is more likely that musically skilled participants only show faster response times and fewer errors. This is already shown in earlier experiments [5].

Summarizing, the experiment described in this paper uses picture categorization as a primary task and mental addition of one-place numbers as a secondary task. The combination of these tasks is known to sufficiently use up mental resources to evoke extra errors in the performance of the primary task. While carrying out these tasks, participants are presented with auditory distracters which bear a certain relation with the information presented visually. Because the auditory distracters are musically coloured, musical skill was considered when selecting subjects. By reaction time analyses as well as error analyses the effects the auditory distracters have on picture categorization are investigated.

METHOD

Participants were 20 students (average age 23) of which 12 were of the male gender. 10 were considered musically skilled because they had been playing a musical instrument for at least 8 years; 4 participants in this group were of the female gender. They were all rewarded with either 10 guilders or 2 participant credits for participating in the experiment.

In the experiment a within-subjects design was used, in which in each of the four conditions the relationship between the picture class and the type of earcon (major or minor mode) was fixed:

- **Congruent condition**: all animal pictures were combined with the major earcon; all non-animal pictures were combined with the minor earcon.
- **Incongruent condition**: all animal pictures were combined with the minor earcon and all non-animal pictures with the major earcon.
- **Neutral condition**: both the animal and non-animal pictures were exhaustively combined with the major as well as the minor earcon.
- **Silent condition**: no sounds were played in conjunction with the pictures.

With this design two different tasks were carried out. In one task the participants only did the categorization, whereas in the other task an extra, mental addition exercise had to be performed. For reasons of counterbalance, participants 1 through 10 carried out the counting task first and participants 11 through 20 started with the no-counting task. Furthermore, to prevent effects of a preferred hand, the position of the 'yes' and 'no'-buttons on the button-box was regularly changed. Finally, the musically skilled participants were equally spread over both task sequences. The participants were instructed to respond as quickly, though accurately, as possible. During the mental addition task it was emphasized that both tasks were equally important and performance on the tasks needed to be equally well. After finishing, the participants were requested to fill in a questionnaire about the sounds they heard during the experiment.

The experiment was carried out on a Macintosh Quadra 840AV. A button-box with voice-key was connected to this computer, as was a pair of headphones through which the earcons were presented. In this experiment the 16 line-drawings used earlier by [4] were modified to include a one-place number in a position close to the center of the drawing. The earcons used were a G chord both in major and minor mode, with a duration of 2500 ms. The alert sound was a single, neutral F-tone.

For data acquisition for a fellow researcher a heart beat measuring device was used [11,12]. The device used was a Polar Vantage NV watch and chestband. In this experiment no further analysis of the heartbeat data was performed however. All participants were notified of the requirement to wear this apparatus in advance of the experiment.

RESULTS

Reaction Times (RT's)

Before starting the statistical analyses, all incorrect responses and null-reactions were filtered out. Of all the remaining responses, responses with latencies smaller than 300 ms and larger than 1500 ms were removed, to make sure that the remaining responses were not reflex-actions or guesses. Therefore, the data considered consisted of all correct responses with $300 \le RT \le 1500$.

The average RT's for both tasks without differentiation on musical skill are shown in Table 1. A repeated measurements analysis showed a significant effect of the extra task (counting) ($F(1,19) = 36.543$, $p = .000$).

In the task with the counting a significant effect of sound was found ($F(1,19) = 21.549$, $p = .000$). This means that in this task, the conditions with sound were significantly slower than the condition without sound. However, between the conditions with sound no significant difference in average RT's was found.

Table 1. Average Reaction Times in ms for the task with the extra counting condition vs. the task without the extra condition.

	RT (ms)	
Condition	**Counting**	**No counting**
Congruent	664	467
Incongruent	708	487
Neutral	684	446
Silent	657	447

For the no-counting task, a significant effect of sound was found as well ($F(1,19) = 5.783$, $p = .027$). Furthermore, a significant difference between the incongruent condition and the other conditions with sound was observed ($F(1,19) = 5.357$, $p = .032$).

Musically skilled vs. unskilled
Differentiating on musical skill again showed an effect of the extra counting task ($F(1,19) = 36.543$, $p = .000$). Table 2 shows the average RT's for all tasks.

Table 2. Average Reaction Times in ms for all tasks differentiated for musical skill.

	RT (ms)			
	Counting		**No counting**	
Condition	**Unskilled**	**Skilled**	**Unskilled**	**Skilled**
Congruent	701	629	470	465
Incongruent	734	681	486	487
Neutral	702	666	433	458
Silent	676	637	442	453

In the task with the extra counting exercise, within the group of musically skilled participants, an effect of sound was found ($F(1,9) = 14.659$, $p = .004$). Between the conditions with sound however, no significant difference in RT's was found. For the musically unskilled group of participants an effect of sound was found as well ($F(1,9) = 7.538$, $p = .023$). No other effects were found however.

For the no-counting task no effects were found to be significant for both the musically skilled and the musically unskilled participants.

Error-analysis
The error analysis showed that 4.1% of the total number of responses were incorrect. Of these errors, 49.75% were made in the counting task. The average number of errors for all tasks is presented in Table 3. A repeated measurements analysis did not show the differences in average number of errors across both tasks to be significant.

Counting vs. no-counting
In the counting task a significant effect of sound was found ($F(1,19) = 16.118$, $p = .001$). Furthermore, the difference between the congruent and the neutral condition, as well as the difference between the incongruent and the neutral condition were found to be significant ($F(1,19) = 9.107$, $p = .007$, $F(1,19) = 4.647$, $p = .044$). The congruent and the incongruent condition were not significantly different.

For the no-counting task the same pattern of effects was found. An effect of sound ($F(1,19) = 17.210$, $p = .001$) was found as well as a difference between both the congruent and neutral condition ($F(1,19) = 4.672$, $p = .044$), and the incongruent and the neutral condition ($F(1,19) = 4.393$, $p = .050$). Again, the congruent and incongruent conditions did not differ in average number of errors.

Table 3. Average number of errors per task and condition

	Average number of errors	
Condition	**Counting**	**No counting**
Congruent	0.35	0.55
Incongruent	0.60	0.65
Neutral	1.25	1.30
Silent	2.75	2.50

Musical skill

Differentiation on musical skill showed no significant differences between musically skilled and musically unskilled participants. Table 4 indicates this.

For the musically skilled participants, doing the counting task, an effect of sound was found again ($F(1,9) = 6.903$, $p = .027$). The congruent condition was significantly different from the neutral condition ($F(1,9) = 8.442$, $p = .017$). The congruent and incongruent condition did not differ significantly. Without the counting task only an effect of sound was found ($F(1,9) = 8.316$, $p = .018$).

Musically unskilled participants carrying out the counting task showed an effect of sound as well ($F(1,9) = 9.596$, $p = .013$), as they did without the counting task ($F(1,9) = 8.548$, $p = .017$). The observed differences between the congruent, incongruent and neutral conditions did not reach the level of significance.

Summary

In both tasks the effect of sound on response time latencies was found again, however, the effect of incongruency was only found in the single task situation. No significant differences between the congruent and neutral condition were found. The differences in RT's for musically skilled and musically unskilled participants did not reach the level of significance. For both musically skilled and musically unskilled participants an effect of sound could only be shown when carrying out the counting task.

The error analysis showed for both tasks an effect of sound. Between conditions with sound, the average number of errors in the congruent condition as well as the incongruent condition was significantly smaller compared to the number of errors in the neutral condition. The differences between the congruent and incongruent condition did not reach significance.

For musically skilled participants carrying out the counting task a significant effect of sound as well as a significant difference in average RT's between the silent and neutral condition was found. None of the observed differences in the single task situation reached the level of significance. For musically unskilled participants only an effect of sound could be shown when carrying out the counting task.

Table 4. Average number of errors for all conditions for all tasks, differentiated on musical skill.

	Average number of errors			
	Counting		**No counting**	
Condition	**Unskilled**	**Skilled**	**Unskilled**	**Skilled**
Congruent	0.400	0.300	0.600	0.500
Incongruent	0.700	0.500	0.600	0.700
Neutral	1.100	1.400	1.500	1.100
Silent	3.300	2.200	2.800	2.200

DISCUSSION

This experiment investigated the effects of auditory distracters on picture categorization with extra cognitive load through the use of a dual-task paradigm. RT analyses show significant slower response latencies when the evoked response of the auditory distracter do not conform to the required response based on the visual stimulus (i.e., an incongruent situation). The data also show a non-significant difference in the neutral condition and congruent condition (evoked auditory response conforms to required visual response). These results confirm the Bussemakers *et al.* results [3,4,5].

Differentiation on musical skill is largely ineffective but for some small scale effects. These latter effects can also been have caused by individual differences between subjects, due to the small group of participants (see for similar results [5]).

Analyses of the incorrect data produced by the dual-task paradigm show no effects of counting, which is an interesting result, because it might seem as if participants do not need additional resources to fulfill the mental addition task. A possible explanation is that participants striving for as few errors as possible, take more time in the dual-task situation. This is clearly indicated by the prolonged response time latencies in the counting task and supports the hypothesis that participants carry out the tasks sequentially [8]. Within each task however, the data show a significant reoccurring pattern of an equal number of errors in the congruent and incongruent condition, a significantly larger amount of errors in the neutral condition and a still larger amount of errors in the silent condition. Again (cf., [5]), no other significant effects are found with musicality as between subjects factor.

Finding confirmation for the earlier results indicates strong consistent effects of auditory distracters on picture categorization [3,4,5]. Addition of auditory distracters to a picture categorization task not only slows down response latencies significantly, it also produces differentiated effects depending on the nature of the auditory distracter in relation to the required response based on the visual stimulus. The consistent effect of sound suggests that the earcons are always processed, but on a subconscious level because participants often are unable to verbalize the differences between the earcons and the relation the earcons have with the pictures (as indicated by the questionnaires). This nonstop processing of the auditory distracters can be directly related to the incapability of the ear to shut out external stimuli like, for example, the eye can by closing the eyelids.

Using the culturally determined connotations of the major and minor mode shows that an auditory distracter in a major mode can significantly slow down the response to a visual stimulus which requires a negative response [6]. However, on the other hand, an auditory distracter in major mode combined with a visual stimulus requiring a positive response does not lead to faster response latencies compared to a situation in which no information can be extracted from the auditory distracter (a neutral situation). These findings might indicate a response-conflict in the incongruent condition in the experiment.

A response-conflict seems an unlikely explanation however, because a response-conflict implies an increase in the average number of errors in the incongruent condition. The error analyses however indicate that no differences exist between the congruent and incongruent condition, because in these conditions the smallest amounts of errors are made, which, for those conditions, did not differ significantly. These error analyses seem to draw a picture in which extra informational resources are processed and always provide some extra information which can be used to decrease the number of errors in conditions in which the extra, relevant, informational resources are present.

Several explanations for the effect exist. One could be that an auditory prime ensures quick responses in participants, because of the alerting function [10]. In this line of reasoning the alerting functionality can cause a state of extra attention in which less errors are made. This explanation does not suffice for two reasons however. First, the auditory distracters in this experiment are not used for attention grabbing purposes in advance of the visual stimulus. For this a neutral F-tone was used and only then the auditory distracter and visual stimulus are presented. Secondly, if the above line of reasoning applies, a difference between the counting and no-counting task should exist, because in the counting task less attentional resources are available due to the secondary task. However, the error analysis only shows a non significant difference between these two tasks.

A more likely explanation is that participants are able to combine the redundant information of the auditory stimuli with the visual stimuli to come to a combined stimulus to which they can respond more accurately, that is, with less errors, than to the visual stimulus itself without the redundant information. Because this is observed not only in the congruent condition, but also in the incongruent condition, this suggests that participants can use the relationship between the auditory distracter and visual stimulus to their advantage, even if the information in the auditory stimuli contradicts the required response to the visual stimulus. The process of combining the auditory and visual stimuli into a unified, singular stimulus can then explain the prolonged response time latencies in conditions with sound.

When applying sound in real-life interfaces the study presented here shows that sound from a performance perspective has no effect, unless it is incongruent with the visual information. In that case performance is hindered by adding auditory information. Applying sound in interfaces therefore still seems to be a case of modest use and careful planning.

REFERENCES

1. Blattner, M.M., Sumikawa, D.A., and Greenberg, R.M., Earcons and Icons: Their Structure and Common Design Principles. Human-Computer Interaction 4(1), 1989, 11-44.
2. Brewster, S.A. Sound in the Interface to a Mobile Computer, in Proceedings of HCII'99 (Munich, Germany, August 1999), 43-47.
3. Bussemakers, M.P., and de Haan, A. Getting in touch with your moods: using sound in interfaces, submitted.
4. Bussemakers, M.P., and de Haan, A. Using earcons and icons in categorisation tasks to improve multimedia interfaces, in Proceedings of ICAD'98 (Glasgow, UK, November 1998) British Computer Society.
5. Bussemakers, M.P., de Haan, A., and Lemmens, P.M.C. The effect of auditory accessory stimuli on picture categorisation; implications for interface design, in Proceedings of HCII'99 (Munich, Germany, August 1999) 436-440.
6. Hevner, K. The mood effects of the major and minor modes in music, in Proceedings of the midwestern psychological association (1933), 584.
7. Logie, R.H. The Seven Ages of Working Memory, in Working Memory and Cognition, Richardson, J.T.E. (Eds.), 1996, 31-65. Oxford University Press, New York, USA.
8. Pashler, H., Johnston, J.C. Attentional Limitations in Dual-task Performance, in Attention, Pashler, H. (Eds.), 1998, 155-189. Psychology Press: Hove, UK.
9. Smith, E.E., Patalano, A.L. and Jonides, J. Alternative strategies of categorization. Cognition, 65, 1998, 167-196.
10. Teichner, W.H. Recent studies of simple reaction time. Psychological Bulletin, 31(2), 1954, 128-149.
11. Van de Ven, J. Touch screen vs. mouse: an experimental comparison using Fitt’s law and mental workload. Available as http://www.shef.ac.uk/pc1jm/hci98cc/hci98cc64.html, HCI’98.
12. Van de Ven, J., de Haan, A. Differences and similarities in the usability of mouse and touch screen. In Proceedings of the international conference on computer-aided ergonomics and safety, Mondelo, P., Mattila, M., Karwowski, W. (Eds.), 1999. Universitat Politècnica de Catalunya: Barcelona.