On Timing and Modality Choice with Local Danger Warnings for Drivers

Yujia Cao
Human Media Interaction
University of Twente
Enschede, The Netherlands
Y.Cao@ewi.utwente.nl

Sandro Castronovo
German Research Center for
Artificial Intelligence (DFKI)
Saarbrücken, Germany
Sandro.Castronovo@dfki.de

Angela Mahr
German Research Center for
Artificial Intelligence (DFKI)
Saarbrücken, Germany
Angela.Mahr@dfki.de

Christian Müller
German Research Center for
Artificial Intelligence (DFKI)
Saarbrücken, Germany
Christian.Mueller@dfki.de

ABSTRACT

We present an experimental study on the effectiveness of five modality variants (speech, text-only, icon-only, two combinations of text and icons) for presenting local danger warnings for drivers. Hereby, we focus on sudden appearing road obstacles within a maximum up-to-date scenario as it is envisaged in Car2Car communication research. The effectiveness is measured by the minimum time necessary for fully interpreting the content. Results show that text-only requires the most time while icon only is perceived the fastest. The two combined versions lie in between. The minimum length for speech is determined by the duration of the utterance, which is longer than perception time of text-only in this case. However, speech could be decoded reliably by nearly all subjects. Results indicate further that a blinking visual cue provided through the periphery visual channel is able to enhance the saliency of visual modalities. Subjective judgements by the subjects furthermore suggest a combined use of visual and auditory modalities.

Categories and Subject Descriptors

H.5.2 [**Information interfaces and presentation**]: User Interfaces, User-centered design

Keywords

automotive, modality choice, timing

1. INTRODUCTION

In-vehicle messages and in particular local danger warnings need to be effective, since the driver has to decode them while being engaged in something else (driving) and because there is typically not excessively much time left to react. The way the information is presented as well as the right timing are therefore crucial factors. One major aspect of the former is choice of the right modality – visual or auditory, which both come with advantages and disadvantages as a number of studies revealed [1, 2, 4, 7]. First of all, there is environmental factors: Visual modalities are superior in delivering information in noisy environments while the performance of auditory modalities is more robust towards variances in lighting conditions. Then, there is consumption of perceptional recourses: Since driving is mainly a visual task, messages that are delivered through the auditory channel can be perceived in parallel with the driving task. In contrast, the perception of visual messages requires to take the eyes off the road. The ability of attracting attention (the level of saliency) has to be taken into consideration as well: Without any additional cue, visual messages are less able to attract attention, especially in a high-load driving condition. When focusing on a busy traffic, drivers might not notice the onset of a new visual message, or they choose to delay attending to it until a moment when they can safely remove their eyes from the road. On the other hand, auditory messages have a preemption effect as to require an immediate perception. When it comes to memory requirements, visual messages allow iterated perception. Auditory information, however, is transient thus might require a repeat function in order to allow recall if it is forgotten.

One might conclude that visual and auditory modalities should complement each other presenting one message. Studies (e.g. [8]) showed that careful combinations outperform the respective inferior modality. However, it has also been found that the result of combining is hardly ever "best of both worlds" [8]. Other studies revealed that people, when occupied by the driving task, tend to only listen to the audio messages and not bother looking at the display on which the same information was presented visually [1, 6]. In addition, a redundant use of both modalities might bear the risk of overloading both perceptual channels or annoying the driver. Therefore, the choice of modality should certainly be done case-by-case. The relevance of the message to the driving task (level of priority) was suggested to be taken into account [2, 8]. When presenting driving-irrelevant messages (such as weather forecast for the coming days), visual modality is more suitable since it is self-paced. The driving task can be better sustained since drivers can take their eyes off the road when the situation allows them to do so. However, if the message to be presented is driving-relevant

AutomotiveUI'09, September 21-22, 2009, Essen, Germany
ACM 978-1-60558-571-0/09/009

(such as warnings of road obstacles), speech should be used in order to allow drivers to obtain a timely awareness of the potential danger ahead.

However, at least one of the drawbacks of visual modality – lack of saliency – might be overcome by means of an additional cue, such as a blinking object located in the peripheral visual field. Peripheral vision is well suited for providing pre-attentive cues because it is sensitive to motion and luminance changes, and it can be picked up in parallel with an on-going foveal vision task [9]. Moreover, there is still the issue of timing. If we are heading towards truly up-to-date warnings, we may not be able to alert the driver well in advance, which would again speak for vision. Imagine a broken vehicle automatically triggering an alert, which is transmitted over ad-hoc car-to-car network and received by a vehicle approaching the place – a scenario that is being investigated in a number of ongoing research projects (e.g. [3]). In a situation like that, we might not have an awful lot of time to generate wordy speech warning messages.

We present an experimental study, which evaluates the choice of modality from a timing perspective. We focus on one aspect of local danger warnings: sudden appearing road obstacles within a maximum up-to-date scenario as described above. The modality choice for presenting this type of messages should assist a fast perception and comprehension of the content of the message. Thereby, the goal of the experiment is to compare the effectiveness of auditory modality (speech in particular) and several enhanced visual modalities. The effectiveness of a modality is measured by the minimum time necessary for fully interpreting the content of the message.

2. EXPERIMENTAL STUDY

2.1 Design/Apparatus

While performing a visual task that required constant attention (simulating driving), subjects were presented with warning messages using different modalities and presentation durations. For each message, the task was to identify whether a repeated version displayed after the offset of the original one was the same or certain details changed. Based on the correctness of the identification performance, we inferred whether or not the presentation duration was sufficient for the subjects to fully perceive the message. The study was performed in a lab room using a PC with a single 20" screen. For the primary task, we chose a "Find the differences" picture puzzle using two versions of a single, very complex photography of a domestic scene. We instructed subjects to perform this task throughout the entire experiment and find a required number of differences. Although the task was interruptible at any time, subjects became very engaged in it because it was very hard to find all differences. The pictures were displayed in a single row on the top left corner of the screen yielding a line of vision that corresponded to looking at the road. Subjects were further instructed to click a button on the screen right below the pictures using the mouse whenever they found a difference. The performance in this task was not analyzed.

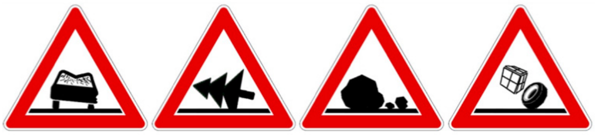


Figure 1: Four types of obstacles were used in the study: break down vehicle, fallen tree, rock, and lost cargo.

Warning messages were displayed on the bottom right corner of the screen respectively played via loudspeakers. A message consisted of three components: 1. type (which kind of obstacle); 2. location (on which lane respectively shoulder); 3. distance (how far the place is ahead). Each visual warning message was preceded by a visual cue (a blinking color bar of the same width as the presentation area and on top of it) and remained on the screen for a certain number of seconds. Two seconds after the message disappeared, it was repeated on the top right corner of the screen together with a choice of three buttons: "same", "different", "not sure". SAME and DIFFERENT cases occurred with a 1:1 ratio and a random order. In the latter case, either type, location, or distance was changed. Since the time interval between the offset of the original message and the repeated message was only 2 seconds, the identification task did not require a long-term memorization of the message. However, it did require subjects to realize what is on the road, where it is and how far it is.

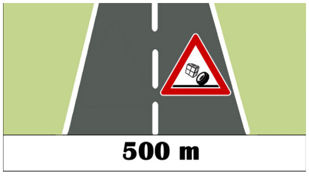
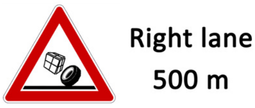
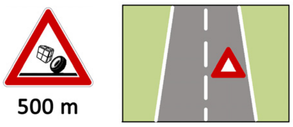
Using a within-subject design, all subjects performed all five presentation styles (see Table 1). The order of the five styles was counterbalanced by a size-5 Latin square. For each visual style, the presentation duration was decreased with a step of 1 second after every three warnings until the subjects started to make errors in identifying the repeated message. We took the minimum presentation duration as a measurement of the effectiveness of this visual presentation style. This includes the time needed to switch the foveal visual attention to the message, perceive the message and understand the meaning. Since in this experiment, subjects always switched their attention immediately when they noticed the blinking motion, our measurement did not include the delay of attentive switch, nor the time needed to prepare an action upon the presented situation. For speech, the length of the utterance was taken as the minimum perception time. Additionally, we surveyed the subjective preferences towards variants of presentation styles.

2.2 Stimuli

Within visual modalities, we further distinguished between textual modality (text, numbers) and graphical modality (icon image), due to the differences in their presentational power. Text is suitable for conveying abstract information, such as the relationships between events. Numbers are suitable for providing precise quantitative understandings of numerical data while images are superior in describing concrete concepts and information of a high specificity nature, such as concrete objects. In general, graphical modalities are more vivid than textual information, thus are likely to receive greater weight during decision making processes. In particular, shapes and colors have great salience to human information processors due to the sharp contrast they are able to create [5].

Four types of obstacles were used: broken vehicle, fallen tree, rock, and lost cargo (see Figure 1). Five modality variants

Table 1: Modality variants used in the experiment.

Variants	example
text only	Lost cargo 500 m right lane
icon only	
mixed 1	
mixed 2	
speech	“Lost cargo in 500 m on the right lane”

were used: speech and four variants of visual presentation (see Table 1). With the visual ones, the distance information was always presented by numbers (e.g. 500 m, 1 km). The obstacle type and the location could be conveyed by either text or icons, resulting in a text only condition, an icon only condition and two mixed conditions (Figure 1). The icons, the wording of text and speech were selected based on a pre-user study with various designs in order to ensure the intuitiveness of the presentation. The textual information (text and speech) were presented in German. Speech was generated using a text-to-speech software.

2.3 Subjects

Ten subjects (2 women and 8 men) voluntarily participated. All of them are German native speakers, between 25 and 45 years old, and working in a technical field (some in speech /dialog-related topics some not). This has to be taken into account when interpreting the results. Interesting correlations could be found especially with subjective preferences.

2.4 Results

2.4.1 Time measurements

The results of the time measurement are illustrated in Figure 2. For visual variants, the center points of the error bars indicate the average of the minimum perception time, and the length of the bars shows the standard error. Text-only required the most time: the average presentation duration which enabled subjects to recall the messages correctly was 3.6 seconds. Icon-only was perceived the fastest. Here, on average only 1.8 seconds were needed to reliably interpret the message and compare it with the subsequent prompt. Not surprisingly, the two mixed version lie in between. However, considering that in both cases only one out of three informational components was replaced by an icon (either type or location), the improvement from 3.6 (text only) to 2.6 for MIXED1 and respectively 2.4 seconds for MIXED2 is remarkable. For the speech condition, the minimum presentation length is determined by the time duration of the utterance, which was 5 seconds on average (Figure 2). The speech messages could be decoded reliably by all subjects, except one who had difficulty to follow the numerical information (distance) in the utterance.

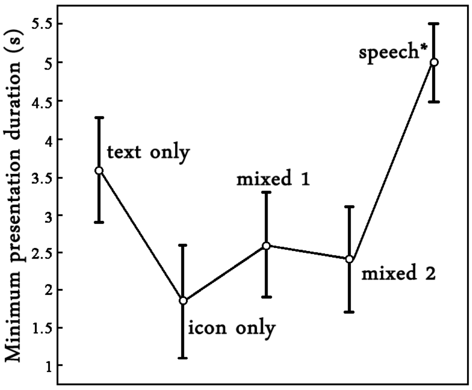


Figure 2: Timing results in seconds for visual modality variants. (*) For speech, the minimum presentation length is determined by the utterance duration.

Table 2: Helmert contrasts between visual variants.

Contrast	Sig.
text vs. icon/mixed1/mixed2	$F(1,9)=30.00$, $p<0.01$
icon vs. mixed1/mixed2	$F(1,9) = 8.65$, $p<0.05$
mixed1 vs. mixed2	$F(1,9) < 1$, $p>0.05$

Repeated-measure ANOVA further showed a significant modality effect on the minimum presentation duration measurement ($F(3, 27)=11.46$, $p<0.001$), indicating that the usage of modality could significantly influence the amount of time needed to perceive and comprehend the same information content. Helmert contrasts (Table 2) further revealed that text-only required a significantly longer perception time compared to the other three; icon-only allowed a significantly faster perception compared to the rest two, and no significant difference was found between the two mixed conditions. Regarding the time needed to decode the message, these results confirmed that text was the least and icon was the most efficient for presenting both obstacle type and location. Although people read the shape of familiar words rather than every single letter, well-designed icons still allow easier perception than text. Being consistent with previous findings, our result showed the representative power of graphical modalities to present concrete concepts such as the obstacle type and location. It also stands in line with the suggestion from [2] that information of higher priority should become more symbolic.

The periphery visual cue was shown to be able to effectively attracts attention. Under the condition that visual presentation gets attention immediately after onset, the time needed to perceive and comprehend the message was much shorter

Table 3: Subjective voting for the best visual variant and visual vs. auditory comparison.

vision variants	text	icon	mixed 1	mixed 2
number of votes	0	8	2	0
vision vs. speech	speech	vision	combination	
number of votes	4	2	4	

than the duration of speech, especially when visual messages were well-designed. Although visual presentation still requires eyes off the road, they might be considered as a good option when a warning message is presented with a short notice (such as shortly before the obstacle). Certainly, speech is suitable when the message is presented long enough ahead and speech should be kept short and precise. Note that our measurement of perception time does not take into account the time needed to prepare an action upon the presented message. However, this time duration is not modality dependent, which means that the difference in the perception time induced by the usage of modality will still be valid even if the reaction time is taken into account.

2.4.2 Subjective judgements

Table 3 summarizes the subjective judgements of the modality variants. When asked to choose the visual variant that they found the easiest to perceive and understand, 8 out of 10 subjects chose ICON ONLY, which is consistent with the minimum duration measurement. They commented that it was time-consuming to read a lot of text. Besides, when interpreting the message, they usually illustrate the text in their mind which requires additional cognitive effort. The reason of disliking the two mixed designs was mainly that the spatial separation of the three information components required longer perception time. The other 2 subjects preferred MIXED1. In contrast to the majority, they explained that they tend to use sub-vocal speech to encode information components into the short-term memory, especially for the location of an obstacle. Therefore it was much more convenient when the location was presented with text. Interestingly, they are the only ones who daily work with language related topics, such as text retrieval and dialog management. These subjective reports indicate that graphical modalities are generally more effective for presenting concrete concepts. However, this conclusion might be moderated by the professional training background of a subject, which might influence the modality used to encode information in the short-term memory.

When asked to compare visual presentations with speech, 4 subjects preferred speech. They stated that speech is more compatible with the on-going visual searching task. Two subjects preferred visual presentations. They said that the visual prime immediately shifted their attention onto the message. However, when they were engaged in the visual searching task, they had a tendency to ignore the speech even though they heard it. The remaining four subjects preferred to be provided with both visual and auditory messages. They stated that, although they listen to the speech, they prefer to have visual presentation as well in case they need to recall details. Moreover, they could choose to look at the visual presentation while the speech output is still ongoing, which is faster for long utterances.

3. CONCLUSIONS

In this user study, we investigated the effectiveness of five modality variants in presenting local danger warning messages for drivers. Hereby, we focussed on sudden appearing road obstacles within a maximum up-to-date scenario as it is envisaged in Car2Car communication research. The effectiveness was measured by the minimum time necessary for fully interpreting the content. Results show that text-only requires the most time while icon only is perceived the fastest. The two combined versions lie in between. The minimum length for speech is determined by the duration of the utterance, which is longer than the perception time of text-only in this case. However, speech could be decoded reliably by nearly all subjects. Result further indicate that the blinking visual cue provided through the periphery visual channel was able to enhance the saliency of visual modalities, thus made them more suitable to present messages of a high priority. When visual messages were attended immediately, the perception time could be much shorter than the duration of speech (5 seconds) for the same information content. This suggests that visual modalities with prime might have advantages over speech when a warning message needs to be presented on a short notice. Speech, however, is certainly suitable when time is sufficient to present the warning. Based on subjective preferences, it might as well be wise to use both visual and auditory modalities. Moreover, our results suggest that spatial integration of information components can reduce the perception time. Generally speaking, our results confirm earlier studies in that it is effective to present concrete information with graphical modalities.

Acknowledgements. This work was funded by the Dutch Ministry of Economic Affairs (project Interactive Collaborative Information Systems, grant nr: BSIK03024) and the German Ministry of Education and Research (project Car-Oriented Multimodal Interface Architecture, grant number 01IW08004).

4. REFERENCES

[1] N. Bernsen and L. Dybkjaer. Exploring natural interaction in the car. In *CLASS Workshop on Natural Interactivity and Intelligent Interactive Information Representation*, pages 75–79, 2001.

[2] A. G. C. Kaufmann, R. Risser and R. Sefelin. Effects of simultaneous multi-modal warnings and traffic information on driver behaviour. In *European conference on human centered design for intelligent transport systems*, pages 33–42, 2008.

[3] S. Eichler, C. Schroth, T. Kosch, and M. Strassberger. Strategies for context-adaptive message dissemination in vehicular ad hoc networks. In *International Workshop on Vehicle-to-Vehicle Communications (V2VCOM).* IEEE Computer Society, 2006.

[4] W. Horry and C. Wickens. Driving and side task performance: The effects of display clutter, separation, and modality. *Human factors*, 46(4):611–624, 2004.

[5] N. Lurie and C. Mason. Visual representation: Implications for decision making. *Journal of Marketing*, 71:160–177, 2007.

[6] M. Moldenhauer and D. McCrickard. Effect of information modality on geographic cognition in car navigation systems. In *IFIP TC.13 Conference on Human-Computer Interaction (INTERACT)*, 2003.

[7] N. Sarter. Multimodal information presentation: Design guidance and research challenges. *International journal of industrial ergonomics*, 36(5):439–445, 2006.

[8] B. Seppelt and C. Wickens. In-vehicle tasks: Effects of modality, driving relevance, and redundancy. Technical Report AHFD-03-16/GM-03-2, GM Corp, 2003.

[9] C. Wickens and J. McCarley. *Applied attention theory.* CRC Press, 2008.