Titel/Title: Device-dependant modality selection for user-interfaces: an empirical study

Autor*innen/Author(s): Christian Elting, Jan Zwickel, Rainer Malaka

Veröffentlichungsversion/Published version: Postprint

Publikationsform/Type of publication: Artikel/Aufsatz

Empfohlene Zitierung/Recommended citation:

Christian Elting, Jan Zwickel, and Rainer Malaka. 2002. Device-dependant modality selection for user-interfaces: an empirical study. In Proceedings of the 7th international conference on Intelligent user interfaces (IUI '02). Association for Computing Machinery, New York, NY, USA, 55–62. https://doi.org/10.1145/502716.502728

Verfügbar unter/Available at:
(wenn vorhanden, bitte den DOI angeben/please provide the DOI if available)

https://doi.org/10.1145/502716.502728

Zusätzliche Informationen/Additional information:

Device-Dependant Modality Selection for User-Interfaces — An Empirical Study

Christian Elting, Jan Zwickel, Rainer Malaka
European Media Laboratory GmbH
Villa Bosch
Schloss-Wolfsbrunnenweg 33
D 69118 Heidelberg, Germany
christian.elting@eml.villa-bosch.de

ABSTRACT

The presentation of information using multiple modalities influences the perception of users, their comfort, and their performance in using a computer-based information system. This paper presents a user study investigating the effects of different output modality-combinations on the effectiveness to transport information and on the user's acceptance of the system. We chose a tourist information system as a test environment and conducted the study on three different devices (PDA, TV set, and desktop computer) to investigate whether the best modality-combination depends on the used device. It turned out that the modality-combination of spoken text in connection with a picture was the most effective regarding recall-performance. This effect was strongest for users working with PDAs, which can be explained by the cognitive load theory. In contrast to this, participants ranked different modality combinations as most appealing, namely those with written text.

neyworns

user studies, user interfaces, multi-modal, device adaptation, cognitive load theory.

INTRODUCTION

The design of intelligent user interfaces becomes increasingly important for software designers and also for the design of consumer electronic devices. This process is due to the fact that software becomes more complex and systems become integrated into a ubiquitous information infrastructure. The paradigm of "one device – one functionality" is over and today we can access mutually any service through any device. A TV set for instance is nowadays a control center for various applications, e.g., video programming, e-commerce, Web-browsing and more. A mobile PDA can be used as a telephone, tourist guide, remote control, calendar, or Web-access device. In the EMBASSI project [2], we are working on multi-modal user interaction in a private household scenario where users access information through different terminal devices such as TV sets, desktop computers, and PDA. The question emerges, how to present the same information on these different device types in order to assist and also to please the user in the best way.

Insights to human processing of multi-modal information can be gained from so-called dual task experiments, in which two different tasks have to be solved at the same time. In general, people perform better when the two tasks are rather different than similar (e.g. [5]). A neuro-psychological account for this could be the separation of different modality-specific areas (e.g. for auditory presented and visually presented texts) in the working memory of the brain (e.g. [5]).

The cognitive load theory of Sweller et al. [6] explains this using Baddeley's model [7] where two separate sub-systems for visual and auditory memory work relatively independent. The load can be reduced when both sub-systems are active compared to processing all information in a single sub-system. Due to this reduced load, more resources are available for processing the information in more depth and thus for storing in long-term memory. Therefore, they propose to use different modalities to present information in order to enhance the learning effect. Of course, this theory only holds when the information presented in different modalities is not redundant.

If the information in different modalities has no new content it only increases the cognitive load [8]. Koroghlanian and Sullivan provide evidences that audio in combination with texts is neither preferred by the users nor effective, at least when the information is redundant [9]. If, however, multiple modalities are used, more memory traces should be available (e.g., memory trace for the information presented auditory, pictorially and visually) even though the information is redundant. This effect would counteract the effect of the higher cognitive load. Furthermore, the theory mentioned above and the following experiments focused on learning-tasks with mostly students as subjects. For non-educational contexts such as entertainment or tourism, we expect differences in the preferred presentation style depending on user experiences. Moreover, these studies have mostly been done using computers and we are

particularly interested in a higher variety of devices including PDAs and TV sets.

In the study presented here, we investigate the effects of display size, device type and style of multi-modal presentation on working memory load, effectiveness for human information processing, and user acceptance. We focused on a task in the tourism domain as a prototypical non-technical domain. This makes a difference to most experiments in cognitive load theory that focus on technical tasks that might be more biased towards using different modalities. For example, we could easily imagine that it is more important to present a technical circuit diagram visually than information about sights.

Another difference to previous studies is the usage of text with longer and more complex sentences not specifically designed for tutorial purposes as done in typical learning experiments (cf. [10]). Short sentences can be easily understood through auditory information only, but longer and more complex sentences might require additional visual presentations with the possibility to re-read them (cf. [11], p. 284 or [12]). Furthermore Tabber et al. [10] showed a modality effect indicating the superiority of auditory over visual information but only for participants who were restricted in the time to process certain information. No modality effect took place when they gathered information at their own pace.

In scenarios concerning private use of devices as in EMBASSI, time restrictions are less realistic than processing information at one's own pace. Besides these aspects of learning and timing that differentiate our study from previous experiments concerned with cognitive load theory, we also wanted to investigate which presentation style users prefer. For private and home applications user acceptance is often more important than effectiveness and in many cases both are, in fact, conflicting design goals.

The remaining of this paper is organized as follows. The following chapter describes the setup of our experiment. We will then present the results and conclude with a discussion. Finally we talk briefly about our future work and summarize the results of the study.

EXPERIMENT

In our experiment, we presented different sights of the city of Heidelberg (Germany) and information about the sights to the participants using different modality combinations (e.g., spoken text with picture, written text with picture). Additionally we used different presentation devices (desktop PC, TV set with remote control, PDA). After presenting the information, we investigated how much of the information was learned by asking the participants to reproduce as much information contained in the texts as possible as well as to reproduce the names of the sights. To assess how appealing the different presentation combinations were for the participants, we asked them to fill out a questionnaire presented after each presentation block at a second computer.

Test persons: We analyzed data of 45 participants. Our sample consisted of 26 females and 19 males, ranging in age from 14 to 58 (mean: 30.3 years). The participants got to know of the study by notices at university locations and a short article in a local newspaper.

Device setup: For the study we used three different devices. We used a desktop PC with a 21'' monitor (HP Ergo 1600) using a 1024x768 resolution and a mouse as input device in the desktop PC environment. For the PDA experiments we used the iPaq pocket PC from Compaq. The iPaq supports a 320x240 screen resolution for graphical output and has a touch-sensitive screen that is used in connection with a plastic pen. Due to the limited disk space of the iPaq (32 MB) we were not able to install the necessary presentation data and software locally. Therefore we used wireless LAN PC cards to export the graphical display of a laptop to the iPaq. Vice versa the touch screen and button input from the iPaq was processed on the laptop. For the TV environment we used a normal TV set with an 800x600 resolution. The graphical output of the TV set came from a laptop, which was connected to the TV set via a SCART cable. As input device we used the same iPaq Pocket PC as in the PDA scenario. The iPaq was connected to the laptop via wireless LAN and worked as a remote control for the TV set. As every presentation originated from the same laptop we achieved the acoustical output of the desktop PC, TV set and PDA by simply using headphones that were plugged into the laptop. Thereby we did not need to implement the transmission of acoustical data via wireless LAN in the TV and PDA settings, respectively.

Stimuli: As stimuli we used 40 sights of Heidelberg with texts varying in difficulty and length. Additionally to each sight name and sight text, a picture (colour or black and white) of the sight could be presented. We implemented three different presentations according to the technical properties of the devices. For that, we modified an existing display manager of our DeepMap system [3]. In the desktop PC environment we used a screen resolution of 1024x768 and were able to present three items at a time. On the TV screen only two items could be displayed in parallel and finally on the PDA only one item could be displayed at a time (fig.1-3). On every device we implemented five multi-modal presentation combinations of text, picture and speech. For speech-output we used the MBrola speech synthesis [1].

The whole setup of the user study including schedules for every user on every device was specified in XML format and read in at presentation time. This served to automatically select the according multi-modality modes as well as the database items (i.e., Heidelberg sights) for a certain participant during the evaluation.

Procedure: Five different multi-modal combinations were used. Written Text only (T), written text with the same text presented auditory (TS), written text with picture (TP), written text with spoken text and pictures (TSP) and spoken

text with picture (SP). We thought this to be a reasonable choice for future applications. Each subject was randomly assigned to a device (Desktop, TV set, PDA) with the constraint that at the end of the experiment the numbers of participants at each device should be equal.

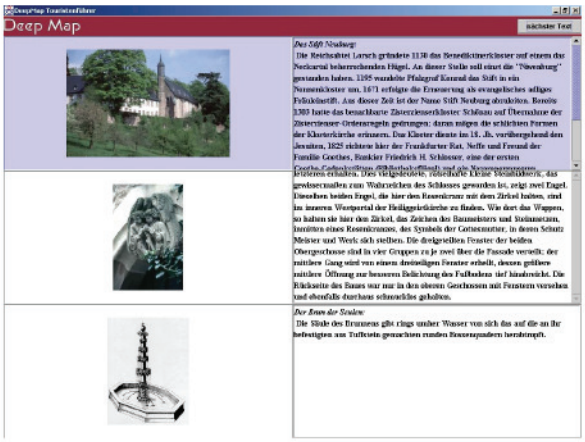

Figure 1: Desktop PC display with text-picture mode.

Figure 2: TV display with text-picture mode.

Figure 3: iPaq display with text-picture mode.

After being assigned to a device the participants were asked to take place in front of a laptop with an electronic questionnaire and answer some general questions concerning their age, sex, experience with computers, PDA and TV sets as well as questions about their general interest in tourist sights.

After this procedure, the subjects were placed in front of the assigned device (desktop PC, TV set or PDA) and were shown eight stimuli in the first modality-combination. Time was taken from onset of each stimulus until the next sight was requested by pressing a key. After that, the subjects returned to the notebook and answered questions about how much they liked the presentation and how much they think that each specific modality in the combination was helpful, being advised not to talk about the content but about the presentation mode. All questions were answered by setting a slider bar to a value ranging from 0 to 100: First we asked the participants how much they were interested in the content of the presentation (consisting of tourist sights). Low values indicated that the participants were not interested in the content of the presentation, whereas high values symbolized high interest. The next question was whether the presentation was overloaded. The participants should use low values if the presentation was not overloaded and high values if they found it too exhausting. A third question concerned how appealing the presentation was. Here low values meant that the presentation was not appealing whereas high values could be used for an appealing presentation. Also, where applicable, the participants should state how helpful graphical text, speech or pictures were for them. Low values should be used if the modality was not regarded as very helpful, high values should be assigned to helpful modalities. Each modality was treated separately.

Then a screen appeared where the participants were asked to enter as much sights-names as they could recall. Additionally, they should state how vivid their impression of the sights was (on a scale from 0 to 100), low values meaning that they could hardly remember anything about the sights, high values indicating that their memory of the sights was very vivid. Depending on the entered names a new screen appeared for each recalled name and asked which information the participants could recall about this sight. The participants could write this information into a free-text field. They were asked to start a new entry field for each remembered information chunk. As examples, information chunks like "built by Franz Meyer" were given.

After that, the participants proceeded to the next part of the study using the same device (desktop, PDA, TV set), but this time with a different modality combination and the next set of sights. Each subject saw all of the modality combinations during the study. The sequence of modality combinations was balanced, i.e., each modality combination was presented at the same place in the sequence (1 to 5) for the same number of times on average for all participants and devices. Furthermore the set of eight sights was coupled with each modality combination for the same number of times over the 15 subjects per device. At each device, exactly the same sequence of sights was used;

meaning that subject 1 in front of the desktop saw the identical presentation as subject 1 in front of the TV set and PDA. The experiment lasted between 1 and 2 hours depending on the individual participant. The participants were rewarded with 25 DM.

RESULTS

The results section is divided into four parts. In the first part general descriptive data are given. In the second part the effect of group affiliation (PDA, desktop, TV) and modality-combination on the time spent in front of the presentation is reported. The third part is concerned with the evaluation of the presentation by the participants in form of acceptance ratings and comments. Finally, in the fourth part, the effect of group and modality-combination on correctly recalled items is reported.

General Descriptive Data

No systematic difference was found between both genders except that females generally rated themselves less competent in handling PDAs, TV sets and desktop PCs than males[1]. Fig.4 displays the reported experience with each device for both genders. Generally, participants rated themselves less experienced with PDA than with desktop.

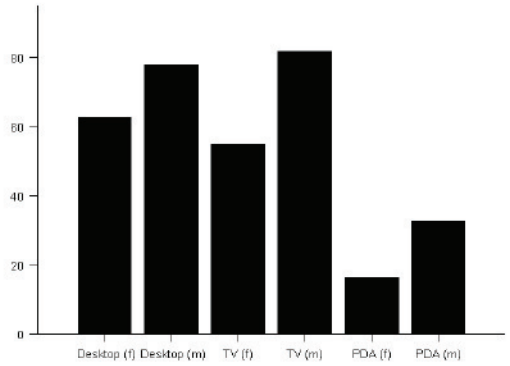

Figure 4: Experience with each device for both genders in % of (f) females and (m) males.

Time Spent at Different Presentations.

Each participant displayed five mean reading times because of the five modality-combinations. There was a significant difference in time spent in front of each device ($F(2, 20)=7.94$, $p<0.0029$), the PDA group needing significantly more time than the two other groups (difference desktop to PDA $t(12)=-3.22$, $p<0.007$). However, no difference between the desktop and TV condition was found ($t(15)=0.77$, $p<0.46$). The interaction between device and modality combination was also significant ($F(8,80)=2.57$, $p<0.02$).

Results of Evaluation by Participants

In comments (given in free text) the participants suggested that a better speech synthesis that would be easier to understand. They recommended that there should be longer pauses between the sentences and that the sentences should be spoken more slowly. Furthermore, participants complained about occasional synthesis errors concerning the production of dates, e.g. "year one thousand nine hundred ninety-nine" instead of "year nineteen hundred ninety nine". In a complete tourist guidance system, they would expect an overview map and the possibility to gather more details on their own demand. The texts should be more structured and if possible, only catchwords should be used. They would like to have more anecdotes related to the sights. Another suggestion was to use rhetoric questions in order to make the presentation more interesting. In conditions, in which no pictures were presented almost half of the participants proposed that pictures should be included. On the other hand in conditions without speech only about ten percent demanded the addition of speech output. Ten percent of the participants mentioned in conditions where text was given in auditory and written form that they want only one form at a time.

There was no significant difference between the best modality-combinations and the second best neither regarding the question about how interesting the presentation content was nor regarding the question whether the presentation was overloaded.

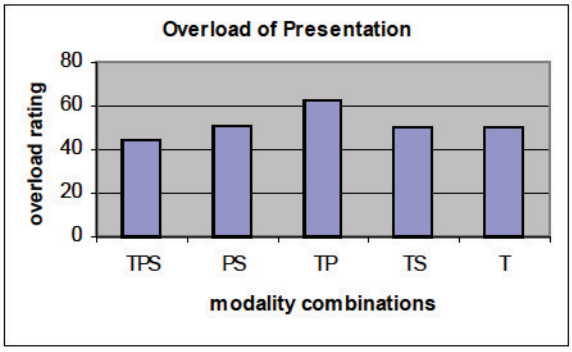

Figure 5: Overload of the presentation modes.

The most interesting modality combination seemed to be the TPS-condition where text was presented written and auditory in combination with a picture of the sight. This combination was also rated quite low in load (fig.5).

TPS was also ranked best concerning presentation appeal in the desktop and TV group (fig.6). Over all groups, TPS was also ranked significantly more appealing than PS ($t(44)=-2.34$, $p<0.02$). Written text was assigned the highest number in condition TP meaning that the participants thought that written text was in this condition a good help. Participants thought that text was less necessary in condition TPS or TS. Written text was assessed as least helpful in the text only condition. Pictures were rated as most useful when they were presented with written texts. Speech seemed to be most helpful in combination with pictures alone.

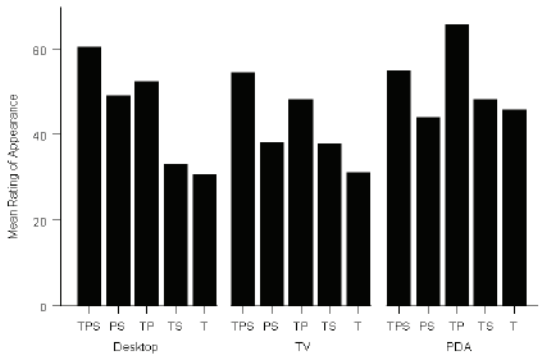

Figure 6: Mean score of presentation appeal.

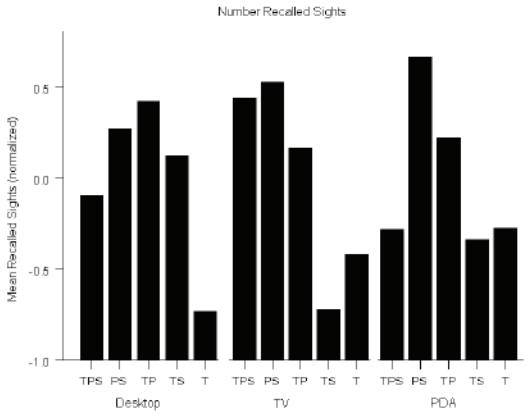

Figure 7: Normalized mean score of recalled items.

Recall Performance

Two measures were taken to evaluate the recall performance: Recall performance of the sight names and the amount of correctly recalled content, where answers were evaluated as correct if they clearly described the correct content of the presented texts. Information that was true but not presented in the text was marked as wrong. For better comparison of modality effects, we also normalized the data of the participants by subtracting the mean of correctly recalled items (for all modality-combinations of the participant) from the number of correctly recalled items (of the modality-combination in question). This value was divided by the standard deviation of the participant over all modality-combinations. This reduces the variance between the participants and allows a relative comparison of the results for different modalities.

Recall of Sight Names: The modality-combination had significant impact on how many sights were recalled ($F(4,176)=8.20$, $p<0.0001$). Fig.7 displays the normalized mean score of recalled items for each modality-combination and for each device. The impact of the devices on the ranking of the modality combinations concerning the number of recalled items was not significant ($F(8, 168)=1.40$, $p<0.20$). The most effective modality combination over all device-groups was PS. In general, the PS results ranked better than the TPS results but the difference was only significant in the PDA group ($t(14)=2.49$, $p<0.03$) and not significant in the others ($t(14)=0.95$, $p<0.36$; $t(14)=0.27$, $p<0.79$). TPS was marginally better in the desktop group ($t(14)=-1.97$, $p<0.07$) and significantly better in the TV group ($t(14)=2.66$, $p<0.02$). There was no significant difference between T and TPS in the PDA condition ($t(14)=0.01$, $p<0.99$). In general, the recall performance decreased continuously during each of the five modality combinations that every participant saw ($F(4,176)=9.18$, $p<0.0001$).

Fig.8 shows the relative recall rate in the PDA group depending on the order of presentation for each modality combination for all modality-combinations over the 5 timeslots.

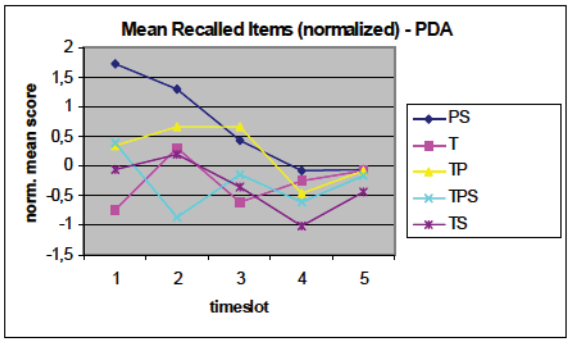

Figure 8: Normalized mean score of recalled items per timeslot on PDA.

Note that each individual received each modality combination only once in the presentation sequence but across the individuals each modality combination occurred at the same frequency at each position in the sequence of presentations. The observed decrease of recall was significant in the desktop and TV group but not in the PDA group (desktop $F(4,56)=5.07$, $p<0.0015$; TV $F(4,56)=3.48$, $p<0.01$; PDA $F(4,56)=1.81$, $p<0.14$).

The decrease in the desktop and TV group was significant for every modality-combination except T (TP: $F(4,25)=5.23$, $p<0.003$; TS: $F(4,25)=4.39$, $p<0.008$; BS: $F(4,25)=3,29$, $p<0.03$; TPS: $F(4,25)=3.68$, $p<0.02$; T: $F(4,25)=0.88$, $p<0.49$). The corresponding figures for the mean scores of recalled items are shown in fig.9 resp. fig.10.

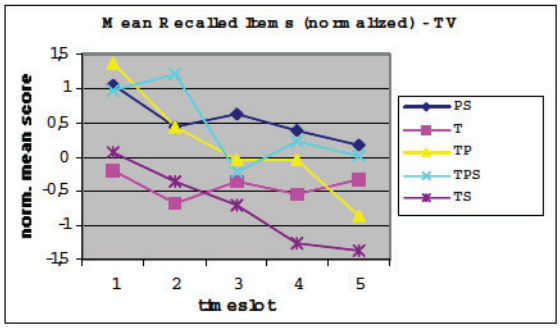

Figure 9: Normalized mean score of recalled items per timeslot on TV.

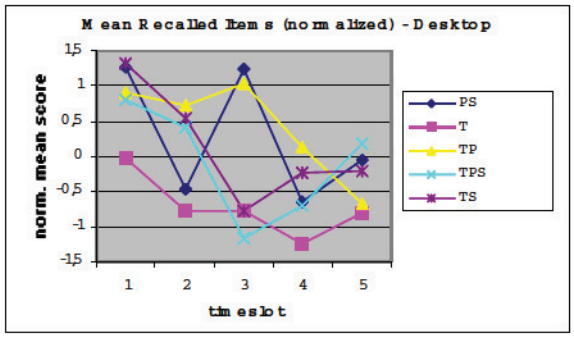

Figure 10: Normalized mean score of recalled items per timeslot on desktop PC.

The question how vivid the participants remembered the sights was answered with higher values in the desktop and TV group than in the PDA group ($t(223)=2.48$, $p<0.01$). The best modality-combination over all groups was PS, T being the worst one. The vivid-ratings were normalized over each participant by subtracting the mean of the participant and dividing by the standard deviation. The difference between PS and T regarding the normalized vividness rating was only significant in the PDA group but not in the other two groups (PDA: $t(14)=4.22$, $p<0.0009$; desktop $t(14)=0.40$, $p<0.70$; TV $t(14)=-0.24$, $p<0.81$). The same pattern was true for the difference between PS and TPS (PDA $t(14)=1.94$, $p<0.07$; desktop $t(14)=0.41$, $p<0.69$; TV $t(14)=-0.37$; $p<0.72$).

Recall of Content: A significant effect of modality-combination on the recall of the general contents of the presentations was found ($F(4,176)=3.02$, $p<0.02$). Over all groups, PS was the best condition, T the worst for correct proposition recall. This difference was significant ($t(44)=2.72$, $p<0.009$). The difference between PS and TS was also significant ($t(44)=2.30$, $p<0.03$). The difference between PS and TPS was not significant over all groups ($t(44)=0.40$, $p>0.69$) but it was marginally significant in the PDA group ($t(14)=1.67$, $p<0.12$).

DISCUSSION

The experimental setup was designed in a way to balance modality-combinations and contents, i.e., each content appeared with each modality and each modality appeared at each position in the sequence of presentations across all subjects. We therefore excluded effects depending on the presentation sequence or the particular content. Between the presentations (each showing eight sights), the questionnaires about the pleasance of the presentations followed by the questionnaire about the sights had to be answered. This ensured that the impression was quite vivid and not disturbed by the decay of memory traces or by recalling sights. Additionally, the short "break" ensured that the participants could not just rehearse the sights but that long-term memory traces must be accessed. According to encoding specifity theory [13] the text modality was preferred a bit by our measure of recall as the recall was done in the text mode, which therefore provided retrieval cues for the text presentation (cf. [14]). This, however, would only be problematic if text presentations were superior to other presentations. Since the results indicate that rather other modality-combinations are better than text, we might only risk that the effect is even higher.

Discussion of Results

General Data: The experiment was done with a representative sample of private users, not being restricted to students or technicians. This is reflected in the range of different ages of our participants. The wide standard deviation of ages also assures that the sample includes many different ages.

Evaluation Data: The comments of the participants suggested to use more pauses between sentences in the speech output, to provide users with overview maps, to use catchwords and to present more anecdotes related to the sights (instead of raw facts). The participants especially pointed out limitations of the speech synthesis. This stresses how important a well-engineered speech synthesis is for a multi-modal user interface.

We were not able to find a significant difference between TPS, PS or TP regarding user interest with TPS being the best-ranked condition. Surprisingly, TPS was not rated as more overloaded than PS and TP. This together with the fact that TPS was ranked more interesting than PS concerning appearance over all groups showed that TPS is a presentation mode that might be of special interest for our future implementations of multi-modal user interfaces. It was expected that text would be rated as more necessary in conditions where the same information was not presented by speech. Surprisingly, text was rated as least important in the text only condition despite the fact that no other information was available. This was interpreted that text without any other information was more difficult to use. It seems that participants needed a picture or at least speech to integrate the information. This interpretation should be taken cautiously and might require further investigations since we found no significant difference between the modality-combinations concerning the evaluation of the usefulness of text. This is also true for the interpretation that pictures are most effective with text only. Additionally

according to the cognitive load theory, two visual presentation modes at the same time are in general not useful and therefore the text with pictures should not be very useful. On the other hand, we did not only focus on performance but also on the individual evaluation by participants and therefore the best modality-combination from the user's point of view could differ from the best combination ranked by user performance. The best evaluation result for the speech modality was found when speech was presented with pictures only. This finding was in accordance with cognitive load theory.

Sight-Name Recall Performance: The results of recall showed a different picture for the PDA group and the other two groups. According to cognitive load theory it is expected that using modalities which can be processed by different loops of the working memory should reduce the load, thereby enhance available capacity for elaboration and storing and therefore should lead to a better performance. According to this theory, PS should be most effective. However, only in the PDA group PS proved to be significantly better than TPS. This result is still in accordance with cognitive load theory. In the PDA group more effort had to be put into the handling of the device and fewer resources were available for the task at hand. Therefore, sparing some resources by using different loops in the working memory through presentation modes should have a bigger effect in the PDA group. Our results supported this assertion. Only in the PDA group the PS condition differed significantly from the TPS condition. In the other groups, PS was most effective even though it did not produce such a big difference, probably due to the generally lower requirements of capacity because of the presentation-mode.

Written text seemed to be a determining variable, especially in the PDA group. While TPS and T differed in the desktop and TV set group, they did not in the PDA group. Probably, more cognitive resources were left in the desktop and TV set group to make use of picture and speech information. The hypothesis of general higher load at the PDA is further supported by the fact that in the desktop and TV set group the timeslot had a big influence on the number of recalled items. If we assume that through time, the number of available resources decreases, it is reasonable to expect that the subjects perform continuously worse. But in the PDA group, participants may have been constantly overstrained by the presentation, and thus time might have not such an impact on recall performance, except in the picture speech condition where no text was presented at all and therefore some capacity was available. A decrease this explicit was neither found in the desktop nor in the TV group as the participants in these groups always had enough cognitive resources left.

Additional support can be found in the vividness ratings of the participants: only in the PDA group the overload produced a significant difference between the different modality combinations again with PS being rated best.

CONCLUSION

In this paper we described the setup and the results of our user study, which was concerned with the effects of multi-modality on recall performance and user acceptance on different devices. In general, the following patterns emerged: the participants found text with picture and speech most appealing. Concerning effectiveness not the most appealing presentation was the most effective one but the presentation expected by cognitive load theory – picture and speech. The biggest impact was found in the modality combinations of the PDA group where reduced recall capacity due to complicated device handling produced the biggest difference between modality combinations.

These results show that cognitive overload is a serious issue in user interface design. A presentation planner has to adapt the presentation to the cognitive requirements of the device used, which is especially true for small mobile devices. However the choice of presentation mode also depends on the system's output goal. When the system wants to present data to the user that is important to be remembered (e.g., a city tour) the most effective presentation mode should be used (in our study picture-speech), which does not cognitively overload the user. When the system simply has to inform the user (e.g., about an interesting sight nearby) the most appealing presentation mode (in our study picture-text-speech) should be used.

We intend to apply those results in the EMBASSI project where we have a large set of output devices and system goals the presentation planner has to deal with. This study served to identify and clarify the problems that occur in multi-device presentation planning.

This study, however, can only highlight some effects on device-dependant modality selection for intelligent user-interfaces. For building actual systems more modality variations should be investigated. For example, we could try to show catchwords besides using auditory and pictorial information or we could allow participants to show more information by clicking buttons. Additionally situations, where spoken and written text delivers non-redundant information could be of interest. Also a more systematical exploration of possible modality combinations could be conducted according to theoretical expectations (e.g. [15]). Moreover, user-parameters, such as experience with particular devices and their influence on multi-modal information perception would be crucial for building adaptive interfaces that automatically select modality combinations according to the user's needs. Furthermore not only the recall performance but also the understanding of the presentation content could be investigated.

ACKNOWLEDGEMENTS

This work was funded by the Klaus Tschira foundation and the German Federal Ministry for Education and Research as part of the EMBASSI project (grant 01 IL 904 D/2). The authors would like to thank Julia Nitschke from Humboldt University, Berlin, for fruitful discussions on the experimental setup.

REFERENCES

[1] The MBROLA Project. Available from http://tcts.fpms.ac.be/synthesis/mbrola.html, 1999.

[2] Herfet, T., Kirste T. and Schnaider, M. "EMBASSI - Multimodal Assistance for Infotainment and Service Infrastructures", EC/NSF Workshop Universal on Accessibility of Ubiquitous Computing: Providing for the Elderly, Alcácer do Sal, Portugal, 2001.

[3] Malaka, R., Zipf, A. "DeepMap: Challenging IT Research in the Framework of a Tourist Information System". In *Proceedings of the ENTER 2000*. Barcelone, Spain, 2000.

[4] Allport, A. B., and Reynolds, P. On the Division of Attention: A Disproof of the Single Channel Hypothesis. *Quarterly Journal of Experimental Psychology, 24*, 225-235.

[5] Smith, E. E., Jonides, J, Koeppe, R. A. Dissociating verbal and spatial working memory using PET. *Cerebral Cortex, 6*, 11-20, 1996.

[6] Sweller, J., van Merrienboer, J. J. G., and Paas, F. G. W. C. Cognitive Architecture and Instructional Design. *Educational Psychology Review, 10*, 251-296, 1998.

[7] Baddeley, A. D., and Logie, R. H. Working Memory: The Multiple-Component Model, in Miyake, A. and Shah, P. (eds.). *Models of working memory: Mechanisms of active maintenance and executive control,* Cambridge University Press, 1999.

[8] Kalyuga, S., Chandler, P. and Sweller, J. Managing Split-attention and Redundancy in Multimedia Instruction. *Applied Cognitive Psychology, 13*, 351-371, 1999.

[9] Koroghlanian, C. M., and Sullivan, H. J. Audio and Text Density in *Computer-Based Instruction, 22,* 217-230, 2000.

[10] Tabbers, H. K., Martens, R. L., and van Merrienboer, J. J. G. The modality effect in multimedia instructions, in Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society (Edinburgh, Scotland), Lawrence Erlbaum Associates Publishers, 1024-1029

[11] Tindall-Ford, S., and Sweller, J. When Two Sensory Modes are Better Than One. *Journal of Experimental Psychology: Applied, 4,* 257-287, 1997.

[12] Sannomiya, M. The Effect of Presentation Modality on Text Memory as a Function of Difficulty Level. *The Japanese Journal of Psychonomic Science, 1,* 85-90, 1982.

[13] Tulving, E, and Osler, S. Effectiveness of retrieval cues in memory for words. *Journal of Experimental Psychology 77*: 593-601.

[14] Elliott, M. L., Geiselman, R. E., and Thomas, D. J. Modality effects in short term recognition memory. *American Journal of Psychology, 94*, 85-104, 1981.

[15] Bernsen, N. O, A Toolbox of Output Modalities – Representing Output Information in Multimodal Interfaces. The Amodeus-II WWW-site, http://www.mrc-apu.cam.ac.uk/ amodeus/, 1995.