Work 41 (2012) 3476-3483
DOI: 10.3233/WOR-2012-0627-3476
IOS Press

The influence of the modality of tele-cooperation on performance and workload

Thomas Alexander, Claudius Pfendler, Jürgen Thun and Michael Kleiber
Research Group Human Factors, Fraunhofer-Institute for Communication, Information Processing, and Ergonomics (FKIE), Neuenahrer Str. 20, D-53343 Wachtberg, Germany

Abstract. Current industrial processes often involve the collaboration of people at distant and remote locations. The technological media for such a tele-cooperation reach from simple email or text-based chatting systems to highly-sophisticated systems for an interactive video-conferencing. But with limited bandwidth the communication between persons at distant locations is often restricted to single modalities. Although this may still be suitable for some tasks, it may result into serious shortcomings and decreased performance with complex tasks like cooperative assembly or maintenance. This is because restricted communication reduces the availability of a common ground, i.e. sharing a common understanding of knowledge, opinions, and goals. The study presented in this paper examines the effect of different communication media on performance of a collaborative assembly task. The results show that tele-cooperation leads to additional verbal communication (AM(direct)=71.1s; AM(video)=145.6s; AM(audio)=204.7s) and, thus, longer times to complete the task (AM (direct)=45.95 min; AM (video)=50.2 min; am AM(audio)=56.16 min). The percentage of relative speech duration also increases significantly. Workload measurement with NASA-TLX did not show any significant differences between cooperation modes. The results allow estimating the effect of reduced communication modalities on time to complete an assembly task. This facilitates a quantification of temporal requirements in time-critical maintenance and repair tasks.

Keywords: Tele-cooperation, tele-communication, multi-modality, assembly task

1. Introduction

With growing complexity of products and a worldwide crosslinking of production chains it is important to provide technological means and media to allow a distributed cooperation. The term tele-cooperation describes such a technology-based cooperation of persons at distant geographic places [21]. It includes many different scientific disciplines including, e.g., business studies, psychology, economics, and computer science. But still the effect of the technological media and infrastructure is crucial as they may limit and restrict communication between the distant persons. Maznewski & Chudoba [18] found that virtual teams can use different media during a decision making process. Simple media (email, fax, short telephone calls) can be used for gaining information, longer telephone calls and conferences for solving problems, and face-to-face meetings for developing new ideas and strategic decisions. According to the model of media richness [3, 17] a medium[*] has to be more comprehensive dependent on the complexity and importance of the communication process. A possible, more technology-oriented approach structures the different media according to textural, auditive and visual output [13]. Each of them has pros and cons so that there is no optimal medium for any type of cooperation. But it can be concluded that audio-visual communication is still not as good as direct cooperation [14].

In addition to the communication medium itself other technical characteristics of the communication infrastructure may also affect tele-cooperation. Above all this relates to *bandwidth*. Bandwidth describes the amount of transferred information. Our own literature research revealed that a minimum of 8 kBit/s is required for audio and 128 kBit/s for video connections, respectively. Applying sophisticated codecs (compressing/decompressing algorithms) may

Corresponding author. Email:
thomas.alexander@fkie.fraunhofer.de

enhance information transfer with limited bandwidth, but introduce additional latency. This is the second important term. *Latency* describes the duration between information input at one location and output at the distant location. Growing latencies, especially latencies larger than 450-700 ms, can restrain communication [11]. Further potentially relevant characteristics are focus of a video camera [7], level of expertise of a participating expert [6], and type of information display [23].

1.1. Tele-cooperation using different modalities for manual assembly tasks

The effects of the technological characteristics grow as soon as manual actions or assembly tasks are considered. In this case, one of the participants of a tele-cooperation acts as an expert instructing the second participant. With this application the communication medium has a major effect on the overall performance.

In general, it is differed between the following modalities and media:

- Audio: The transfer of speech requires only a small amount of technological performance. Speech can be transferred with minimum bandwidth and also allows using both hands for manual works.
- Audio and text: In addition to speech, text messages can be sent. This has benefits for sending complex information for later references. However, it restricts manual work.
- Video and audio: The participants of a tele-cooperation can share their view while talking to each other. This has benefits for cooperative manual work when the object can be visualized.
- Video, audio and text: The participants can communicate with each other by speech and text, and a visual connection between them is available.

An important advantage of video communication is a shared visual space of both participants of the working area. In this case both can support each other [15, 10]. Situational awareness and a common mental model are supported, thus facilitating common problem solving. A reason for this is a more efficient communication about objects within the working area, because of, e.g., the possibility of simply pointing to objects.

Fussel et al. [6] investigated the effect of the *communication media* on tele-cooperation. The authors compared side-by-side collaboration, i.e. direct cooperation, to audio only and audio/video cooperation. The results showed that the assembly task was performed faster and more precisely with a direct cooperation compared to tele-cooperation. Although the audio/video cooperation also provided video images, a shared visual space was not realized because of a mismatch between the camera position and orientation of both participants. However, the authors did use a within subject design with different types of tasks and communication media. This might have introduced confounding leading to larger unexplained variance and, thus, to larger errors.

Another important factor which critically influences the performance of tele-cooperation is the so called *grounding* of both participants [2]. Grounding describes an interactive process between communicating persons, during which both gain knowledge about the specific goals, opinions, positions etc. of each other. Additional actions of one another, e.g., nodding the head, might be interpreted correctly. Grounding is clearly dependent on the communication media and available technology. However, the influence of different visual information sources on grounding is still under discussion. Although capturing the participants' mimic might support a common grounding, it might still be irrelevant to the task and, thus, the overall performance.

1.2. Measures of performance for tele-cooperation

There are different measures of performance for tele-cooperation. The first one is the *time to complete the task*. It is obvious that performance decreases with reduced time to complete the task. A second important variable is the *probability or frequency* of errors. It might be increased with audio communication because of missing visual control by the expert. These variables are task-dependent.

The second group of variables is referring to the communication of both participants. One is the *duration of communication*. The second is the *number of questions*. Both can increase with limited means for communication because of misunderstandings and additional explanations.

The third group of variables relates to more complex constructs: Workload and situational awareness.

Workload is defined according to Hart and Staveland "as the cost incurred by a human operator to achieve a particular level of performance" [12]. Therefore, this construct is important to be considered when estimating effectiveness and efficiency of a technical system. Workload can be measured by subjective (rating scales, questionnaires, etc.) or objective measures (physiological measures, performance measures, etc.) [16, 19]. The implementation of the tele-cooperation system is assumed to have significant effect on workload, because additional communication, unclear instructions etc. will increase the overall difficulty of the task and, thus, also increase workload.

The second construct, *situational awareness (SA)*, is another important dependent variable. SA has been defined as "the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [5]. Factors with negative effect on tele-cooperation might also hinder SA because of limited understanding of the meaning of actions and shortcomings in projecting the effect of actions into the future. SA is usually measured by means of questionnaires. An example is SAGAT (Situation Awareness Global Assessment Technique) [5], which relates to the three areas of SA. Other questionnaires relate to task demands on attentional resources, supply of attentional resources and understanding of the situation [24]. However, as the future status of the task used in our experiment could not be predicted by the helper, SA was not measured.

2. Method

The goal of the following experiment was to investigate the influence of the modality of tele-cooperation on performance and workload. The application field that was of critical importance was a manual assembly task in which a tele-cooperation participant assembled a LEGO® model and was assisted by a remote expert. There were two kinds of tele-cooperation between the participant and the expert – video and audio – which were compared to direct (face-to-face) cooperation. The last experimental condition served as a baseline for a comparison. By changing the modality of cooperation visual perception is continuously diminished. At the same time verbal communication increases to compensate this deficit. This can be seen, e.g., for video-based cooperation. In this case vision for the expert might be restricted by a fixed camera whereas there is no visual restriction in direct cooperation. With direct cooperation the participant can point to a critical part of an object and the expert can focus this part. This part might be hidden under the video condition. The most restrictions appear for the audio condition, where extensive and complex verbal explanations might be necessary so that the expert can locate a critical part. This might also increase workload. Consequently, it is expected that tele-cooperation does not only increases communication time but also search time to locate a critical part of an object.

This is also why ecological interface tries to facilitate problem solving by visualization in order to reduce workload. As a consequence reserve capacity can be invested into other tasks [20].

2.1. Hypothesis

Based on these considerations the following hypotheses were derived:

1. Task completion time*:* Time to complete a task is increased from direct cooperation to video- and audio-cooperation. This can be attributed to longer communication and search times.

2. Communication time: Time for verbal communication is increased from direct cooperation to video- and audio-cooperation, caused by reduced visual perception.

*3. Workload***:** Workload is increased from direct cooperation to video- and audio-cooperation. The progressive constraint of visual perception must be compensated by extensive and complex verbal communication thus increasing workload.

2.2. Participants

30 male participants of the Fraunhofer FKIE volunteered to participate in the study. They were between 20-35 years .

2.3. Apparatus

Experimental task: There were several preconditions for the task:

- The task had a mechanical-technical background, so that the results could be generalized, e.g., to a car repair task.
- The participants should not have any prior knowledge on the task
- Task difficulty should be high enough, so that the participants required support of an expert
- Completion time should be at least half an hour

These preconditions were fulfilled by a LEGO® kit of a model excavator (No. 8047) consisting of 252 parts which had to be assembled manually by the participants (Figure 1). For this purpose the participants used the original LEGO® working plan.

Fig. 1: LEGO® model for manual assembly

Matching test: In order to realize three matched groups for the three experimental conditions the participants performed a pretest in which they had to assemble the vehicle chassis of the excavator (the first 16 steps of 56 in the working plan of the model). A high correlation between the completion times of the matching test and the main task for assembling of the model excavator was an important precondition for comparability of the three experimental groups.

NASA-TLX: NASA-TLX [12] was applied for workload measurement in the cooperation task. This subjective rating method consists of six subscales (mental, physical and temporal demands, own performance, effort and frustration) which are weighted and combined into an overall workload score.

Tele-cooperation system: The experimental tele-cooperation system for the participant and the expert included of a TFT-monitor, a headset and a web camera. The camera was focused on the desktop surfaces of both participants. It captured the model excavator of the participant and presented it on the monitor of the expert. For the audio- and video-cooperation between the participants the video conference program EKIGA [1] was used. The audio signal was recorded and conversation times of the participants were measured.

2.4. Procedure

All 30 participants conducted the pretest at first. On the basis of the completion times the subjects were rank ordered. To match the three experimental groups the rank order was divided into groups of three participants and these three members of each group were randomly assigned to the three experimental conditions.

After that the participants had to assemble the remaining parts of the excavator. The parts of the model were arranged in a predetermined way in a case, so that all participants started the experiment under the same preconditions. As it was not completely sure that the participants needed assistance from the expert during the task, four steps in the working plan of 40 steps were omitted so that questions from the participant were nearly unavoidable.

For constant experimental conditions the expert had to consider four rules which were derived from Fussel et al. [6]:

- each question has to be answerd by the expert,
- each question must be answered as good as possible,
- if the participant does not ask for a minute, the expert must offer assistance,
- in direct cooperation the expert does not intervene manually.

In *direct cooperation* the participant and the expert cooperate face-to-face. As in all other experimental conditions the participant used the working plan of the model, which was displayed on his monitor. According to the instructions the participant should proceed step by step as fast as possible without making mistakes.

In *video-cooperation* the participant and the expert were situated in separate rooms and the expert could see the participant working at the model on his monitor. The participant was able to watch the desktop surface of the expert and the expert could support the participant by gestures and both could communicate verbally

In *audi-cooperation* the participant and the expert were also in different rooms. Only verbal communication between the participant and the expert was possible.

It was also assumed that learning of the expert could play a considerable role in the experiment. E. g., at the end of the experiment he might perhaps detect mistakes of the participant faster and verbalize solutions more precisely. To compensate for this learning effect the sequence of experimental conditions was randomized in the experimental design. So, learning could not favor any experimental condition.

2.5. Variables

The *independent variable* of the experiment was the modality of cooperation with the levels of direct cooperation, video- and audio-cooperation.

The *dependent variables* of the experiment were: Task completion time which the participant needed to assemble the model. Furthermore, verbal communication was recorded and from that conversation time of the participant was derived as absolute and relative score in relation to time for task completion. After task completion the participant rated workload of the task with a computerized version of NASA-TLX.

2.6. Statistical analysis

As the experiment used matched groups based on pretest results all variables were analyzed with a single factor within subjects ANOVA [4, 22]. In case of significant main effects subsequent comparisons of means were made with Scheffé tests [4].

3. Results

3.1. Correlation between pretest and main task completion time

The correlation between the pretest and the main task completion time proved to be highly significant ($r=0.809$, $p<0.001$) so that it was concluded that the three experimental groups were comparable with respect to the task relevant variables.

3.2. Task completion time

Task completion time for assembling the model is illustrated in figure 2. The model is assembled fastest in direct cooperation (am: 45.9 min). With video-cooperation the participants required 50.2 min and with audio-cooperation 56.2 min. on the average. There are significant differences between cooperation modalities ($F=3.930$, $p=0.038$, $df=2/18$). Audio-cooperation shows a significantly longer completion time than direct cooperation ($p<0.05$).

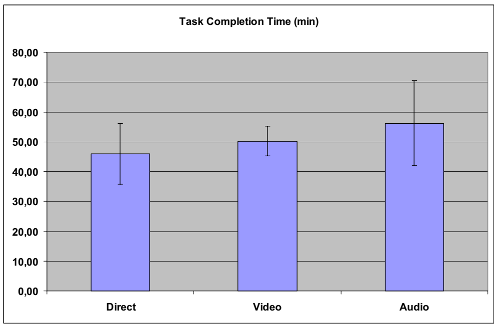

Fig. 2: Task completion times for assembling the model (means and sd)

3.3. Conversation time

Figure 3 shows the conversation times for the different modalities. In comparison to direct cooperation (am: 71.1 s) average conversation time approximately doubles with video-cooperation (am: 145.6 s) and triples with audio-cooperation (mean 204.7 s). There are significant differences between the modalities of cooperation ($F=10.436$, $p<0.001$, $df=2/18$). Conversation time is significantly longer in audio-cooperation than in direct cooperation ($p<0.01$).

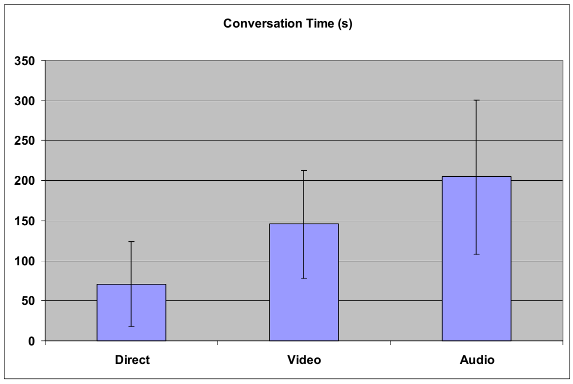

Fig. 3: Conversation times when assembling the model (means and sd)

3.4. Relative conversation time

Conversation time in relation to completion time is shown in figure 4.

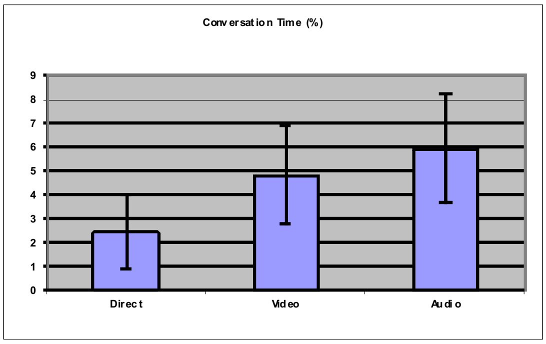

Fig. 4: Relative conversation time when assembling the model (means and sd)

The percentage of conversation time at the task completion time increases from direct cooperation (2.43 %) to video-cooperation (4.78 %) and audio-cooperation (6.04 %). The differences are significant (F=9.239, p=0.002, df=2/18). In comparison to direct cooperation relative conversation time is significantly longer with video-cooperation (p<0.05) and audio-cooperation (p<0.01).

3.5. Workload

Figure 5 shows the NASA-TLX weighted workload score as a function of modality of cooperation.

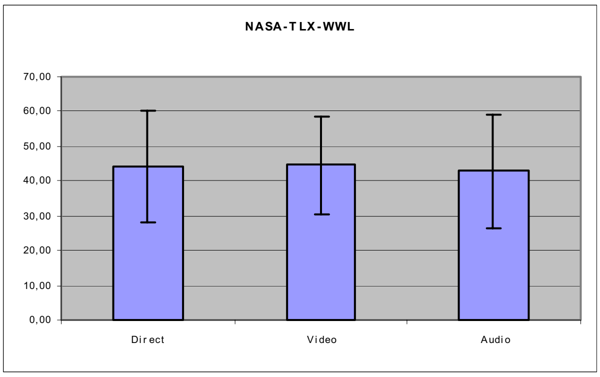

Fig. 5: NASA-TLX weighted workload score as a function of modality of cooperation (mean and sd)

With all cooperation modalities an average workload level can be observed. The individual ratings do not differ significantly (F=0.067, p=0.935, df=2/18) from each other. Furthermore, the NASA-TLX subscale scores do not differ significantly between the experimental conditions.

4. Interpretation of results

Current industrial processes often require the collaboration of people at distant and remote locations. Tele-cooperation is an approach to overcome associated problem by supporting a technician by an expert at a remote location. Our study has examined the effect of direct cooperation and two tele-cooperation modalities on performance and workload in a collaborative assembly task.

In the experiment a participant and an expert assembled a model excavator in a face-to-face cooperation, in a video- and an audio-cooperation task. In order to achieve comparable groups for these three experimental conditions we used matched groups. The high correlation between the matching test and the main task shows that the three experimental groups were successfully matched.

Hypothesis (1) of an increment of completion time from direct cooperation and video-cooperation to audio-cooperation was partially supported as the latter modality of cooperation took significantly more time than the first. However, there was no significant difference in completion time between video- and audio-cooperation showing that the experts' vision on the task object obviously does not play such an important role as expected in supporting the participant.

Hypothesis (2) of an increment in conversation time from direct cooperation and video-cooperation to audio-cooperation was also partially supported as the latter modality of cooperation took significantly more time than the first. Furthermore, relative conversation time also supported this hypothesis. In addition, video-cooperation required more conversation time than direct cooperation. But again, there was no significant difference in conversation time between both modalities of tele-cooperation.

The results also show that direct cooperation is definitively the most efficient modality of cooperation. With video- and audio-cooperation at least longer relative conversation times are found than with direct cooperation. But both modalities for tele-cooperation do not differ significantly from each other with respect to performance. Audio-cooperation shows no significant increment in completion- and conversation time than video-cooperation. The additional video information affords considerable more bandwidth, but does not result in a relevant improvement in tele-cooperation. Therefore one should consider thoroughly if video-cooperation has any advantages in a task.

The relatively small increment in conversation time between all cooperation modalities (from 71.1 s to 204.7 s) could not explain the considerably higher increment in task completion time (from 45.9 min. to 56.16 min.). This discrepancy could be explained by the observation that the participants often tried to solve problems without the support of the expert, although this was contrary to the participants' instructions. When the expert observed an error in direct cooperation, he could intervene. In tele-cooperation the expert's vision on the task object was restricted (video-cooperation) or totally concealed (audio-cooperation) so that he could not always give immediate recommendations. As a consequence, the participants often continued their task until they detected their mistakes. The return to a previous step in the working plan was often time consuming. So, the objective data and the observations show the influence of common grounding, which is continuously reduced from direct cooperation to video-cooperation and audio-cooperation.

Hypothesis (3) addressing workload was not supported by the results. When reducing the visual information between the participant and the expert a workload increment could not be observed. An approximately average workload level was found for all modalities of cooperation, without significant differences between the modalities. This is an important aspect in respect to tele-cooperation, as the lack of visual information in audio-cooperation obviously did not result in a higher workload level in comparison to video-cooperation. It was expected that complex explanations of both participants, e.g., in respect to kind of parts would increase workload as the kit included many parts which could be easily confounded. Furthermore, location of parts had to be verbalized. But obviously visual information does not always seem to be relevant for all kinds of tele-cooperation in respect to performance and workload. On the other side it must also be considered that the result might depend on the camera focus. In the present experiment the camera focus corresponded to that in an experiment of Fussell [7] which had proven to be best in their investigation.

Another explanation for the workload results could be that there were relatively few sequences with "gaps" in the working plan in comparison to the sequences without gaps. The latter were comparable under all three experimental conditions and might have dominated the workload ratings, so that comparable rating results were found in all experimental conditions.

In summary, it is obvious that direct cooperation is the best alternative. Audio-cooperation increased completion and conversation times considerably in comparison to direct cooperation. In contrast to the hypotheses video-cooperation and audio-cooperation did not differ significantly. The results have to be interpreted on the background of the relatively short time delay in tele-cooperation (0.3 s) in the present experiment. An increment in the time delay, which seems to be quite realistic, could even increase completion and conversation times.

5. Conclusion and Future Work

The results of the experiment show that even with a direct video of the task object the relative conversation time was still longer than during direct cooperation. This is a clear indication that more than just standard tele-cooperation equipment is required in order to bridge the gap between a technician and a supporting remote expert. The results differentiate between the expert requiring a better understanding of the remote workplace or the mechanic needing better instructions to fulfil the task. Both options promise to support the cooperation so that the performance increases.

To help the mechanic understand the given instructions it can be beneficial to integrate them into the real view of the workplace i.e. through augmenting reality (AR). For example, by showing an animation of the desired action it is very easy for the mechanic to understand what part he needs to pick up and what operation he needs to perform.

The expert can also benefit from an animated view of the next steps or an artificial 3D view of the maintenance object. By using the 3D data aquired for AR by the mechanic's support system it is even possible to recreate the mechanic's view [25].

We are investigating both of the possibilities outlined above and will investigate their impact on tele-maintenance performance.

Acknowledgement

We gratefully acknowledge funding of the project by the German Federal Office of Defense Technology and Procurement (BWB), Team T5.2.

References

[1] Ekiga(2011):Internetpage:http://download.chip.eu/de/Ekiga-fuer-Windows_4388419.html. Visited. 08/2011

[2] Clark, H.H. & Brennan, S.E. (1991): Grounding in Communication. In L.B. Resnick, R.M. Levine & S.D. Teasley (Eds.).

Perspectives on socially shared cognition. Pp. 127-149. Washington DC: APA.
[3] Daft, R.L., Lengel, R.H. (1986): Organizational information requirements, media richness and structural design. Management Science, 32.
[4] Eimer, E. (1978): Varianzanalyse. Kohlhammer, Stuttgart
[5] Endsley, M. R. (1989): A methodology for the objective measurement of pilot situation awareness. In: Situational Awareness in aerospace operations. AGARD Conference Proceedings No. 478. Neuilly Sur Seine, France.
[6] Fussell, S. R., Kraut, R. E., & Siegel, J. (2000): Coordination of communication: Effects of shared visual context on collaborative work. *Proceedings of CSCW 2000* (pp. 21-30). NY: ACM Press. Download Preprint (PDF)
[7] Fussell, S. R., Setlock, L. D., Kraut, R. E. (2003): Effects of Head-Mounted and Scene-Oriented Video Systems on Remote Collaboration on Physical Tasks. CHI 2003, Fort Lauderdale, USA. Pp. 513-520.
[8] Fussell, S. R., Setlock, L. D., Parker & E. M. & Yang, J. (2003): Assessing the Value of a Cursor Pointing Device for Remote Collaboration on Physical Tasks. CHI 2003, Fort Lauderdale, USA. Pp. 788-789.
[9] Fussell, S. R., Setlock, L. D., Parker & E. M. (2003) : Where do Helpers Look? Gaze Targets During Collaborative Physical Tasks. CHI 2003, Fort Lauderdale, USA. Pp. 768-769.
[10] Gergle, D. (2005) : The Value of Shared Visual Space for Collaborative Physical Tasks. CHI 2005, Portland, Oregon, USA. Pp. 1116-1117.
[11] Gergle, D., Kraut, R.E., Fussell, S.R. (2006): The Impact of Delayed Visual Feedback on Collaborative Performance. CHI 2006, Montreal, Canada. Pp. 1303-1304.
[12] Hart, S.G. and Staveland, L.E. (1988): Development of NASA-TLX (Task Load Index): Results of experimental and theoretical research. In P.A. Hancock and N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: North-Holland.
[13] Hertel, G. Geiste, S., Konradt, U. (2005): Managing virtual teams: A review of current empirical research. Human Resource Management Review, 15.
[14] Hilt, Volker & Geyer, Werner (1997): A Model for Collaborative Services in Distributed Learning Environments. [Hrsg.] -. LNCS. London, UK: Springer, 1997, Bd. 1309, S. 364 - 37.
[15] Kraut, R. E., Gergle, D., Fussell, S.R. (2002): The Use of Visual Information in Shared Visual Spaces: Informing the Development of Virtual Co-Presence. CSCW 02, New Orleans, USA.
[16] Lysaght, R.J., Hill, S.G., Dick, A.O., Plamondon, B.D., Linton, P.M., Wierwille, W.W., Zaklad, A.L., Bittner, A.C., and Wherry, R.J. (1989): *Operator workload: Comprehensive review and evaluation of operator workload methodologies* (Tech. Report 851). Fort Bliss, TX: U.S. Army Research Institute, Field Unit.
[17] Maruping, L.M. & Agarwal, R. (2004): Managing team interpersonal process through technology: A task-technology fit perspective. Journal of Applied Psychology, 89.
[18] Maznevski, M.L. & Chudoba, K.M. (2000): Bridging Space over time: global virtual team dynamics and effectiveness. Organization Science, 11.
[19] Pfendler, C., Pitrella, F. D. und Wiegand, D. (1995): Messung der Beanspruchung bei der Systembewertung. FAT, Wachtberg-Werthhoven, Bericht Nr. 115.
[20] Pfendler, C. und Thun, J. (2010): Geografische Orientierung mit egozentrischen und geozentrischen Karten auf einem Head-Mounted-Display und einem Personal Digital Assistant. FKIE-Bericht Nr.192, Fraunhofer FKIE, Wachtberg-Werthhoven.
[21] Reichswald, R., Möslein, K., Sachenbacher, H. & Englberger, H. (2000): Telekooperation, verteilte Arbeits- und Organisationsformen. Berlin: Springer.
[22] SPSS 15.0 (2006): SPSS Inc. Headquarters, 233 S. Wacker Drive, Chicago, Illinois 60606
[23] Tang, A. ,Owen, Ch., Biocca, F. & Mou, W. (2003): Comparative Effectiveness of Augmented Reality in Object Assembly. CHI 2003, Fort Lauderdale, USA. Pp. 73-80.
[24] Taylor, R. M. (1989): Situational Awareness Rating Technique (SART): The development of a tool for aircrew systems design. In: Situational Awareness in aerospace operations. AGARD Conference Proceedings No. 478. Neuillly Sur Seine, France.
[25] Kleiber, M. & Alexander, T. (2011): Evaluation of a mobile AR tele-maintenance system . Proceedings of the 14th International Conference on Human-Computer Interaction. LNCS 6768, Orlando, USA.