A Study on Visual, Auditory, and Haptic Feedback for Assembly Tasks

Bernd Petzold
bernd.petzold@iwb.tum.de
Institute for Machine Tools and
Industrial Management
Technische Universitaet Muenchen
Boltzmannstrasse 15
D-85748 Garching, Germany

Michael F. Zaeh
Institute for Machine Tools and
Industrial Management
Technische Universitaet Muenchen

Berthold Faerber
Barbara Deml
Human Factors Institute
University of the Armed Forces,
Munich

Hans Egermeier
Johannes Schilp
Stella Clarke
Institute for Machine Tools and
Industrial Management
Technische Universitaet Muenchen

Abstract

Telepresent tasks involve removal of the human operator from an immediate worl ing area and relocation to a remote environment that offers the operator all neces sary control features. In this remote location, the operator must be provided with adequate feedback information such that the task at hand can be effectively executed. This research explores the effectiveness of various feedback methods. More specifically, graphical feedback in the form of video streamed images is compared against rendered 3D models, the overall effectiveness of haptic feedback is analyze and the influences of sensory augmentation and sensory substitution are examined This study involved 48 participants, each of whom executed a simple clockwork assembly task under various feedback mechanisms. The results support the use of 3D models as opposed to live video streams for graphical presentation, utilization of haptic feedback (which was found to significantly enhance operation effectiveness), and the use of sensory augmentation and substitution under specific circumstances.

1 Introduction

Microsystems technology deals with manipulated objects that are less than 1 mm in size or for which an assembly accuracy of less than 200 μm is required. This kind of miniaturization is becoming increasingly important for automotive, medical, and telecommunication components. Today, the assembly process of these components is faced with at least two major problems: th motor demand on the human operator is high due to the requirement of acci rate results, and, to avoid any contamination, the human operator has to wor in a clean room. Both factors are a threat to good working conditions and cai be avoided through the use of telepresent technology. In telepresent operations, the assembly process is performed by a slave manipulator controlled by human operator located in an external environment. This remote assembly changes several aspects of the working environment.

First, the visual presentation as well as the operator's actions must be scalec such that the user gets an impression of being in a "normally" sized environment. The scaling is not the only parameter that has to be adapted at the ope ator end. The visual image can either be a camera-recorded real video stream or a rendered 3D model. The video setup is easier to apply because no mode need to be created. Furthermore, no synchronization is necessary as there are no reality drifts as observed in virtual models. In a 3D model, distracting infc

Presence, Vol. 13, No. 1, February 2004, 16–21

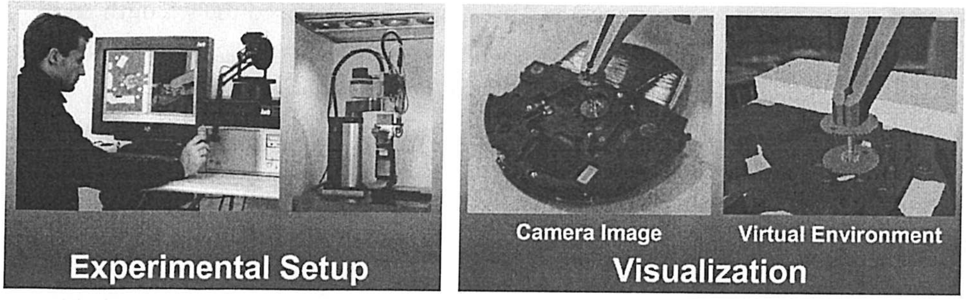

Figure 1. *Setup of the pilot experiment.*

mation can be suppressed, and the viewpoint can be changed easily. To determine the optimal visualization system, the following experiment was conducted. At the Institute for Machine Tools and Industrial Management (hereafter referred to as the *iwb*), an experimental telepresence system for micro-assembly was realized. A Bosch SR6 Turboscara robot was used at the teleoperator end. A micrograsper was attached to the robot to pick up microsized objects in a tweezer-like manner. Between the grasper and the main body of the robot, a six-DOF-force/torque sensor was fixed to measure the contact forces. At the operator end, a PHANToM 1.5–6 DOF was used to apply the forces to the user and to control the teleoperator. For the visualization, two cameras were installed. One camera showed a bird's-eye view of the assembly scene, and the other camera displayed a side view. The grayscale video stream from the two cameras was shown directly on the operator's screen. The 3D model for the virtual environment was generated using a 3D CAD program that is usually applied for the development of complex new products. The scene was displayed in a viewer program based on the graphic API Open Inventor, with which it was possible to display two views of the same scene simultaneously. The communication between the operator and teleoperator utilized a CORBA communication framework developed at the *iwb* (Reinhart, Anton, Ehrenstrasser, Patron, & Petzold, 2001).

The experimental task consisted of mounting an hour-wheel onto the corresponding minute-wheel of a wristwatch clockwork with a positioning accuracy of 20 μm (Figure 1). In total, the task was executed by 60 participants who were divided into two groups, each using the different visualization systems. To ensure comparable conditions, the viewpoints of the virtual environment were fixed such that only one bird's-eye view and one side view were available. Each trial started with the hour-wheel already grasped by the gripper, and every participant had to repeat the assembly operation three times, while the contact forces, the position of the tool center point, and the time to complete were monitored. Of special interest are the results concerning completion times and average forces: visualizing the scenario with a virtual model makes the assembly process not only faster, but also decreases the contact forces compared to a video image. In detail, the group using the camera needed 50% more time in the first attempt and 30% more time in the third. Also, the average contact forces during the operation were 10 to 25% higher than those for the virtual visualization group (Reinhart, Anton, Ehrenstrasser, & Petzold, 2001). Based on these results, it was decided to use virtual visualization in future experiments.

Scaling the scene and choosing the appropriate presentation matches only the visual feedback channel; force feedback, which is also an essential part of a real-world assembly, is not yet considered. Of course, the most straightforward way to display forces is through a haptic feedback device. However, in practice this is not always easy to administer. In such cases, force feedback can be substituted by, or augmented with, other senses as discussed in Section 2.

2 Sensory Substitution and Augmentation

Humans are able to gather the same information with more than one modality. Blind people, for example, use their tactile sense to substitute visual information, whereas deaf people substitute auditive information with their visual sense. In the context of man-machine systems, a common example of sensory substitution is a park distance control system for cars. Here, the remaining distance is provided through an audio alert signal. In a similar manner, it is possible to translate force information into visual or audio format. Hereby both the intensity and the direction of a force can be expressed by varying the loudness or the pitch of a tone or by adapting a visual stimulus in length, brightness, or color. Both cross-modal displays have proven to be successful for virtual environment systems: Bergamasco (1992) showed the effectiveness of a symbolic visual arrow. Massimino and Sheridan (1994) as well as Edwards (2000) demonstrated that audio signals are a suitable substitute for force feedback. Lécuyer et al. (2002) studied a combined audiovisual environment. They found that, although the audio stimulus was accepted as a warning signal, it did not improve performance. In addition, they noted that the applied visual stimulus seemed to be too complex and was not helpful either. In conclusion, although sensory substitution is comprehensible to the operator, there exists the additional cognitive burden of transforming visual or audio information into the force domain.

Another way of benefiting from intermodal coding is through sensory augmentation. Here, haptic feedback is not substituted but rather supported by force information. Richard and Coiffet (1995), for example, displayed haptic feedback together with an auditive force stimulus. They found that this enhancement yielded better results than presenting either the force or the audio information alone.

In this study, audio as well as visual force information are examined in the context of sensory substitution and augmentation. The audio output occurred each time the operator caused a collision and sounded like two metallic surfaces colliding. The visual output was designed as a bar graph that displayed the strength and direction of collision forces. Both displays were realized in six different combinations:

- *substitution:* (1) visual + auditive (2) visual (3) auditive
- *augmentation:* (1) haptic + visual + auditive (2) haptic + visual (3) haptic + auditive

The authors assume that both outputs are suitable fo displaying force information (Bergamasco, 1992; Massimino & Sheridan, 1994). Although bar graphs are a convenient way of providing detailed information and i would seem natural to provide visual feedback in this way, interpreting a bar graph is less intuitive than reacting to a collision sound (Lécuyer et al., 2002). For this reason, the authors suppose that, when haptic feedback is available, the visual information is neglected as being too abstract and that optimum results will be achieved with auditive augmentation only (Richard & Coiffet, 1995). However, when haptic feedback is missing, it is likely that the operators will use all information availabl (even when it is time consuming). Hence, it is assumed that visual information augmented with audio information will yield optimal results in the substitution condition:

- *Assumption 1:* Sensory substitution will be most successful when visual and auditive force cues are presented together.
- *Assumption 2:* Auditive augmentation can especiall improve the effect of haptic feedback.

3 Experimental Design

The three types of sensory information—haptic, auditive, and visual force information—were systematically varied to study sensory substitution as well as sensory augmentation (Table 1).

Based on the results of the first experiment, the visu scene was displayed as a virtual scenario. Therefore, the operator end could be isolated from the remaining system. Running the experiment as a simulation has the advantage that unintentional disturbances that may be caused by sensors or data transmission can be avoided.

Table 1. *Experimental Design: Each Condition (Cell) was Presented to Six Participants*

Auditive	Visual	Haptic	
		with	without
with	with	6	6
	without	6	6
without	with	6	6
	without	6	6

Total number of subjects: 48.

Figure 2. *Hardware setup for the experiments.*

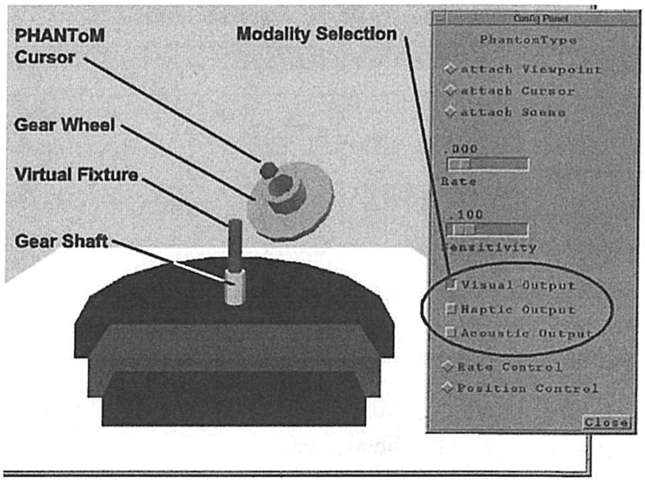

Figure 3. *Experimental procedure.*

The underlying simulation system, Ve^2 (Virtual engineering environment), which was developed at *iwb,* is based on the commercial toolkits WTK for visual rendering and Solid 3.1 for collision detection. The visual rendering and collision detection are computed in parallel to provide a fast system response and therefore high-quality visual and haptic feedback. As the haptic control device, a PHANToM is applied, which offers force feedback of up to 8.5 N s (Figure 2).

A pick-and-place task serves as the experimental scenario. The subjects have to grasp a simplified virtual gear wheel and attach it to a virtual gear shaft. If the PHANToM cursor is in contact with the gear wheel, it can be picked up with a button click on the PHANToM stylus button. The subjects are then able to move the grasped object in six degrees of freedom to place it as required. Furthermore, a virtual fixture is offered, which functions as a permeable anchor (and therefore does not provide feedback). This additional helping aid increases depth perception and facilitates navigation in the three-dimensional virtual space (Figure 3).

Many of these features were explained and demonstrated to the subjects. The operator (i.e., subject) learned how to move the device, grasp the object, and profit from the virtual fixture. However, the participants were not explicitly told whether they would experience a certain sensory feedback or not. Subsequent participants could run one test trial in which task performance was not timed. Afterward, the task had to be performed as fast as possible, and the time was measured from the first movement of the device until a certain defined target area was reached. Finally, a presence questionnaire was completed, which consisted of three subscales (Scheuchenpflug, 2001). The first factor addressed the sensation of being a part of the remote environment and can be called "spatial presence." The second factor refers to the interaction with the control device and the display and could be termed "quality of interface." The third factor is composed of items addressing "emotional involvement."

Table 2. *ANOVA Results for the Presence Variable*

ANOVA	*F*-statistic	Significance level
haptic	4.795*	.03
auditive	0.358	.70
visual	0.293	.59
auditive × visual	0.004	.95
haptic × auditive	0.053	.95
haptic × visual	0.320	.57
haptic × auditive × visual	3.749	.06

A significant effect (<.05) is observed only for the factor haptic as marked by the asterisk.

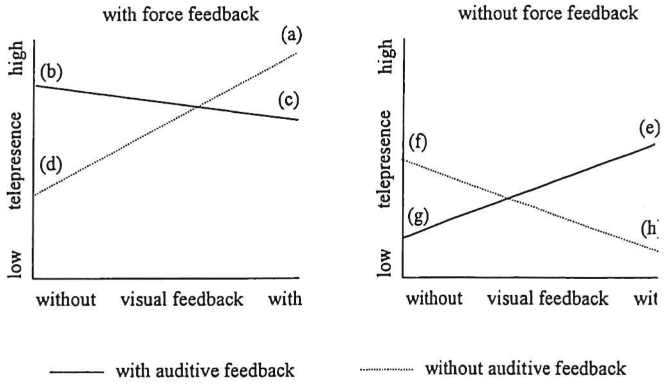

Figure 4. *ANOVA interaction diagram for sensory augmentation (left) and sensory substitution (right). Both figures have to be interpreted separately. The interaction effect is visible as both lines do not run parallel.*

4 Results

First, the correlation between the completion times and the subjective presence scores was regarded. A two-sided Neyman-Pearson test revealed a significant negative correlation between both measures. In general, this means that participants with a high sensation of presence managed to complete the task quicker.

Next, the effects of audio, visual, and haptic feedback were examined through an analysis of variance (ANOVA). Table 2 displays the values for the presence scale. Although the results for the variable "time-to-complete" are similar, a significant effect (significance level < .05) could be detected for only the presence variable. Here, a significant main effect for the factor "haptic" (significance level .03) is observed, which indicates that the participants feel more present when haptic feedback is available.

Because the three-way interaction factor "haptic × auditive × visual" almost achieved a significance factor of .05, its interaction diagram is analyzed (Figure 4).

The left figure displays the results under the haptic feedback condition and therefore can be interpreted in terms of sensory augmentation. The highest sensation of presence is obtained when force feedback is either enhanced by a visual bar graph or (a) an auditive stimulus. (b) However, sensation of presence decreases when both augmentation displays are available, (c) although augmentation always seems to be superior compared to limiting feedback to the haptic channel (d).

The right figure displays the results when haptic feedback is not present and can therefore be interpreted in terms of sensory substitution. When haptic feedback is absent, the participants derive most benefit from a combination of both auditive and visual cues (e), although i has to be mentioned that this sensory substitution is no much better than the absence of both visual and audio force feedback (f). It is, however, important to remember that the user still receives "image feedback" in the form of the rendered 3D model. An interesting result is that either the auditive (g) or the visual substitution (h) displayed alone seem to be more confusing than helpful

In summary, in this study a high sensation of presenc goes along with good performance. In addition, haptic feedback contributes significantly to the sensation of presence. In situations in which haptic feedback is not possible, audio and visual feedback together provide a suitable substitute. Although this result supports the author's Assumption 1, the effect of this substitution is not large and only little better than visualizing the scen solely. Especially interesting is that both cues seem to interact and compete each other, while when displayed alone each of them is experienced as a distractor and no as an aid.

Besides this, haptic feedback should be enhanced by either a visual or an auditive augmentation. In contrast to the authors' Assumption 2, more benefit is derived by the detailed visual display than the auditive display. This means that the visual cue provides further information, which cannot be felt immediately. For this reason, it is likely to assume that the actual haptic feedback requires further improvement to become more distinctive. Finally, augmentation seems to be limited to one substitution channel and cannot be applied according to the notion of "the more, the better."

5 Conclusions

In this study, the application of telepresence technology to micro-assembly tasks is examined whereby special emphasis is placed on the design of the master site. The first experiment studies whether the working environment of the slave robot should be presented as a live camera image or as a virtual image to the operator. The second experiment concentrates on the haptic feedback channel. Should haptic feedback be substituted by or augmented with other modalities? Therefore, occurring collision forces are expressed as auditive as well as visual output. As an experimental scenario, a pick-and-place task was chosen. In terms of interface design, the most important results can be summarized as follows.

- A virtual presentation outperforms a realistic camera view when time-to-complete and applied contact forces are regarded. This is because the virtual image limits presentation to the most relevant elements for the task.
- The application of haptic feedback is recommended. Although haptic feedback is less important regarding time-to-complete, it seems to be an essential contribution to the sensation of presence. Furthermore, the best outcome is achieved when haptic feedback is enhanced by either a visual or an auditive augmentation.
- In cases in which haptic feedback cannot be provided, it is best to apply auditive and visual force information together, although the effect of this substitution is only little better than limiting the presentation to pure visualization.

To focus interface design on more than the design of feedback, various input devices and bimanual control were studied in another recent experiment. Future work will encompass a comparison of telepresent and real-world micro-assembly.

Acknowledgments

The presented work is funded as part of the SFB 453 Collaborative Research Center "High-Fidelity Telepresence and Teleaction" of the DFG (Deutsche Forschungsgemeinschaft).

References

Bergamasco, M. (1992). The GLAD-IN-ART Project. The GLAD-IN-ART project. *Proceedings IMAGINA '92,* II, 7–14.

Edwards, G. W. (2000). *Performance and usability of force feedback and auditory substitutions in a virtual environment manipulation task.* Unpublished master's thesis, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Lécuyer, A., Megard, C., Burkhardt, J.-M., Lim, T., Coquillart, S., Coiffet, P., et al. (2002). The effect of haptic, visual and auditory feedback on an insertion task on a 2-screen work-bench. Proceedings of the Immersive Projection Technology (IPT) Symposium.

Massimino, M. J., & Sheridan, T. B. (1994). Sensory substitution for force feedback. *Presence. Teleoperators and Virtual Environments, 2*(4), 344–352.

Reinhart, G., Anton, O., Ehrenstrasser, M., Patron, C., & Petzold, B. (2001). Framework for teleoperated microassembly systems. *Proceedings of SPIE,* 86–96.

Reinhart, G., Anton, O., Ehrenstrasser, M., & Petzold, B. (2001). Telepresent microassembly at a glance. *Workshop Proceedings: Advances in Interactive Multimodal Telepresence Systems,* 21–31.

Richard, P., & Coiffet, P. (1995). Human perceptual issues in virtual environments: Sensory substitution and information redundancy. *Proceedings of the IEEE International Workshop on Robot and Human Communication,* 301–306.

Scheuchenpflug, R. (2001). Measuring presence in virtual environments. *Proceedings of the 9th International Conference on Human-Computer Interaction,* 56–58.

PRESENCE

TELEOPERATORS AND VIRTUAL ENVIRONMENTS

OLUME 13 / NUMBER 1 / FEBRUARY 2004

MIT PRESS