

Automation in Surgery: The Impact of Navigated-Control Assistance on the Performance, Workload and Situation Awareness of Surgeons

Maria Luz¹, Stefan Mueller², Gero Strauss², Andreas Dietz², Juergen Meixensberger² and Dietrich Manzey¹

¹Berlin Institute of Technology, Berlin, Germany

²ICCAS Innovation Center Computer Assisted Surgery

The present study investigates performance consequences of a new approach of automated support for surgeons. "Navigated-Control" (NC) represents an advancement of image-guided navigation that does not only support the surgeon in navigating through a patient's anatomy, but also can stop the surgical device if it comes too close to risk structures which need to be protected to ensure patient safety. It is explored, how NC affects different aspects of surgical outcome, workload and stress, and situation awareness. Fourteen advanced students of medicine performed a simulated Mastoidectomy with and without NC support. The results reveal that NC support can reduce both, the risk of intra-operative injuries, as well as the physiological stress level of surgeons. However, "cost effects" emerged with respect to subjective workload, and a reduced spare capacity compared to unsupported surgeries. These latter effects do not seem to be related to the principle of NC but technical constraints of current implementations.

INTRODUCTION

Image-guided navigation (IGN) represents a rapidly advancing technology in providing computer-based support for surgical procedures. It has specifically been developed in order to support the surgeons' spatial orientation within and navigation through the anatomy of a patient. Based on a (preoperative) digital computer or magnetic resonance tomography dataset of the patient's anatomy, a three-dimensional (3D) camera, a specific registration device and sophisticated tracking technology, these systems can automatically display the location of the surgical instrument in relation to the specific anatomical structure of a patient on a screen. In order to get this information, the surgeon just needs to bring a specific navigation pointer to the same location as the surgical instrument. First systems of this kind were developed in the 1980s and have now been in clinical use for more than a decade. The most prominent areas where these systems have been applied, thus far, include neurosurgery, functional endoscopic sinus surgery and other skull base surgeries (Koele, Stammberger, Lackner & Reittner, 2002).

Technically, the currently available pointer-based IGN tools can be regarded as a kind of automation where the cognitive task of navigation is no longer performed by the surgeon alone but shared with a machine (Manzey, Strauss, Trantakis, Roettger, Bahner-Heyne, Lueth, Dietz & Meixensberger, 2009). Referred to the framework model of human-automation interaction proposed by Parasuraman, Sheridan & Wickens (2000) this corresponds to a partial automation of information acquisition and analysis functions. A first human factors assessment has revealed strength and weaknesses of this approach in terms of human performance consequences (Manzey, Roettger, Bahner-Heyne, Schulze-Kissing, Dietz, Meixensberger & Strauss, 2009). Beneficial effects were found to be related to an improved surgical performance as well as an increase of situation awareness with respect to a correct orientation within the surgical site. However, negative side-effects were identified with respect to an increase of time-pressure and mental workload. These latter effects seem to be mainly related to some limitations of the

currently available pointer-based IGN tools: the necessity of instrument change for navigation, changes in the surgeon's line-of-sight axis, and limited length of use of the navigation information.

The principle of "Navigated-Control" (NC) represents a recent advancement in this area. In contrast to the pointer-based tools it enables for a continuous tracking of the surgical instrument (Strauss, Koulechov, Hofer, Dittrich, Grunert, Moeckel, Muller, Korb, Trantakis, Schulz, Meixensberger, Dietz & Lueth, 2007). Thus the surgeon does not need to handle a separate instrument in order to get the navigation information which is continuously available. Furthermore and even more important, the automation provides autonomous support for protecting risk structures in the anatomical area where the surgeon is working. Based on a preoperative segmentation of workspace (excluding the critical structures) in the digital set of tomographical data, the system continuously assesses the distance between the surgical instrument (e.g. drill) and the workspace border and automatically stops the instrument in case this border gets reached. This latter function represents a significant enlargement of range of automation. In contrast to the currently available pointer-based IGN systems, the NC approach does no longer represent just a kind of information automation that supports information acquisition and analysis functions but a tool that also provides decision support (e.g. where to work and where to stop). This impacts the kind of human performance consequences associated with this technology. Beneficial effects might be expected with respect to patient safety. i.e. a reduced risk of injuring nerves or other important anatomical structures which are present and need to be protected. In addition, the availability of such technology might make it possible also for relatively inexperienced surgeons to perform complex surgeries. However, also new issues might arise concerning the impact of this new technology on trust, reliance and situation awareness. For example, surgeons who trust the system too much might tend to neglect their own monitoring during the surgery and may lose awareness for anatomical specialities that are important to notice and consider.

The current study presents the first approach to address these questions experimentally. The model used for this research is the performance of a Mastoidectomy by a surgeon who has not got much experience with this surgical intervention so far. A Mastoidectomy is a particular risky intervention at the petrosal bone behind the ear. The task is to remove as much as possible of the infected portion of the mastoid by means of a mill (medical term: trephine) without injuring relevant risk structures present in this particular anatomic area. Risk structures to be protected include the *dura mater*, the *sigmoid sinus*, the *auditory ossicles* and lateral *semicircular canal*, and – most important – the *nervus facialis*. In addition to investigate general effects of the availability of NC assistance on performance, workload and situation awareness, another objective of the experiment is to explore whether the availability of NC assistance will enable comparatively inexperienced surgeons to perform this complex intervention without compromising patient safety.

METHOD

Participants

17 advanced students of medicine and physicians who just had finished their exam participated in the experiment. Three participants had to be excluded from data analysis because of technical problems during the experiment or because they needed an extremely long time for the simulated surgery (> 3hrs). Thus, the results presented are based on 14 participants (5 male, 9 female; mean age: 26 years). On average the participants had first clinical experiences from a 4.5 weeks clerkship (range: 0-12 weeks), and 11 of the participants already had first surgical experiences from an internship. None of them had any experience in performing a Mastoidectomy. Participants were paid € 25 as compensation for participation in the study.

Task

The primary task consisted of a simulated Mastoidectomy. A special simulation device was used for this task (see below). Participants were instructed to perform this task as good as possible without injuring any of the risk structures which are present in this anatomical area. They further were instructed to work according to a given sequence of steps, i.e. (1) break through the *corticalis*, (2) expressing of the *dura mater*, (3) expressing of the *sigmoid sinus* (4) lancing of the *antrum*, (5) expressing of the posterior wall of the auditory channel, (6) expressing of the *sinus-dura* angle, and (7) expressing of the *nervus facialis*.

In addition a secondary-task had to be performed. This task represented a simple probe-reaction time task. The participants had to react on an acoustic signal as quick as possible by pressing a foot pedal. The signal was presented randomly with a mean inter-stimulus-interval of 90 seconds and a standard deviation of this interval of 5.5 seconds.

Apparatus

An artificial skull with an exchangeable *petrosal bone* module was used for simulation. The skull and the mastoid modules were made with a 3D printer on the basis of CT images of real patients. All aspects of the mastoid modules were designed to correspond as closely as possible to real mastoid anatomies. This included typical levels of pneumatization of the bone material, as well as a simulation of all risk structures which are represented in this anatomical area, i.e. *dura mater*, *sigmoid sinus*, *nervus facialis*, *auditory ossicles* and lateral *semicircular canal*. All risk structures but the *auditory ossicles* were additionally equipped with integrated sensors which made it possible to automatically log any intra-surgical injuries of these structures. Two kinds of mastoid anatomies were used, one with normal pneumatization and normal location of all risk structures, and another one, also normally pneumatized but with an exteriorized *sigmoid sinus*.

The NC-system (KARL STORZ Surgical Cockpit ENT) used for navigated-controlled interventions consisted of a 3-D camera system, a trephine equipped with sensors needed for instrument tracking, a navigation display, and the control unit. In order to use this system for the simulated mastoidectomy in the experimental trials, the risk structures of the artificial petrosal bone were segmented-out in the digital CT data set by one of the co-authors (S.M.). This corresponds to the usual clinical preparation needed for use of navigated-control assistance (Strauss et al., 2007).

Physiological data (ECG, respiration) were recorded with a mobile recording device (NeXus-10, version 2008a; Mind Media B.V. Netherlands) which automatically generated estimates of mean heart rate and respiration rate for defined periods. Identification of R-waves in the ECG data for analysis of heart-rate variability (HRV) was accomplished by using *Nevrokard HRV File Preparation* software (Version 9.2.2) and afterwards corrected manually, if necessary. The resulting time-series of inter-beat intervals was further analysed by *Carspan for Windows* software (Version 0.0.1.26) developed by Ben Mulder (University of Groningen).

For blood pressure recording a standard anaesthesia device was used which was available in the operating theatre where the experiment was conducted.

Design

Each participant performed the Mastoidectomy twice, one time with and one time without NC assistance. The sequence of the two experimental runs was balanced across participants.

Procedure

The experiment consisted of two training sessions and two experimental runs. Before the first training session all participants were provided a script which familiarized them with the anatomical characteristics of the petrosal bone, the critical risk structures in this area, the different steps of a Mastoidectomy, and the relevant surgical instruments. The

actual training consisted of two practice runs (distributed across two different days) where the participants had to perform the Mastoidectomy on normally pneumatized petrosal bone modules with standard anatomy. These runs had to be performed manually, i.e. without NC assistance. They were performed to familiarize the participants with the use of the microscope, the use of the trephine (including different milling heads), and the identification of the different risk structures which need to be expressed during the intervention without injuring them. In addition, they ensured that the participants acquired sufficient basic skills to perform the experimental trials without the risk that basic problems of navigation and device handling would interfere with the effects of the experimental treatments.

The two experimental runs were distributed across two different days. They took place in the endoscopic operating theatre of the ORL University Hospital in Leipzig. After the participants had dressed the scrubs, they had to read and to sign the data privacy policy (only first day), fill in a demographical form (only first day) and then to read the instruction. Thereafter each participant had the opportunity to familiarise with the microscope and the use of the suction apparatus for removing the dust produced during milling, as well as to perform the trephine registration procedure. In case of NC assisted runs a short introduction in the navigation system and its functions followed which was provided and demonstrated by one of the experimenters.

Before starting the simulated surgery, several baseline measures were collected. These included a measure of single-task response times in the secondary task (only first day), a measure of blood pressure, and the recording of a five minute baseline of ECG and respiration data. Participants were instructed to rest and relax during this period.

The simulated surgery itself was performed in sitting position. The participants sat in front of the surgery table on which the artificial skull was placed. The surgical microscope was placed directly in front of them and adjusted to the individual needs. House spoons, trephine and milling heads, kidney basin with water for the cooling of milling heads, as well as the registration tool were placed on the right side of the skull, where the patients' body usually would be situated. On the left side the head of the suction apparatus was placed. The trephine was controlled by a foot pedal which was located underneath the surgery table. Left of this pedal another pedal was located which was used to register responses to the secondary task. The camera of the navigation system and the navigation screen were placed in front of the participants. In manual conditions the navigation screen was out of their sight.

The Mastoidectomy had to be performed according to the standard protocol. However, the anatomy of the modules was varied by using modules with an exteriorized *sigmoid sinus*. This variation was introduced in order to simulate the normal inter-individual variety of anatomies and to elevate the risk of a serious injury of the simulated patient. After the lancing of the *antrum*, the participants were asked to interrupt and to turn to the experimenter. Immediately after this interruption a situation awareness questionnaire was presented to the participants. This procedure equalled the Situation Awareness Global Assessment Technique (Endsley, 2000). In

addition to the administration of the situation awareness questionnaire, a measure of blood pressure was taken. Thereafter the surgery was continued until the participants indicated that the intervention was completed. Upon this, another situation awareness assessment was conducted. In addition, subjective workload ratings were collected and post-measurements of all physiological parameters were performed before the participant left the operating theatre.

Two experimenters were present during the entire experimental run, one of them responsible for data recording and handling of the experimental hardware, and the second one for communicating with the participants, for supporting them to fill in the SA questionnaires and for keeping a log of important events.

Dependent Measures

Surgical performance. Four different performance measures were defined: (1) the time needed to complete the Mastoidectomy, defined as the sum of time (in minutes) from the first start of the trephine until the lancing of the *antrum* (where the surgery got interrupted), and from the restart of milling after the interruption until the indication of the participant that the surgery had been completed, (2) the number of injuries of risk structures which were automatically registered by the simulation software; (3) an expert rating of the overall quality of the surgical outcome which was provided by an experienced surgeon who assessed all modules in a blind procedure on a 5-point rating scale ranging from "very good" to "very bad", (4) an expert rating about whether or not seven possible complications might be expected to arise from the given Mastoidectomy; for each complication, ratings had to be provided on a 4-point-scale (sure no, rather no, rather yes, sure yes).

Workload. Three different kinds of measures were used to assess different aspects of workload: (1) Subjective workload was assessed with the NASA Task Load Index (Nasa TLX, Hart & Staveland, 1988), (2) Measurements of physiological effort included mean heart rate, mean heart rate variability (defined as mean power in the range of 0.07-0.14 Hz), mean respiration rate, and systolic and diastolic blood pressure; (3) Secondary-task response time.

Situation Awareness. The Situation Awareness Global Assessment Technique (SAGAT, Endsley, 2000) was used to assess SA. Upon the interruption after the lancing of the *antrum* the participants were required to answer six different questions which focused on an estimation of the distances between the last position of the milling head and several defined anatomical structures, a report on the already accomplished tasks, and specific aspects of the anatomy (i.e. location of the *dura mater* and *sigmoid sinus*; level of pneumatization). Three indicators of SA were derived from the answers: (1) the rank correlation between the estimated and real distances calculated across the six risk structures; this measure reflects how accurate participants were aware about their current position in the 3D space of anatomy (0-1); (2) number of correctly remembered steps (0-7); (3) number of correct answers about the anatomical characteristics (0-3).

RESULTS

Performance

A significant difference between the two experimental conditions was found with respect to the time needed to complete the Mastoidectomy. Participants needed less time to perform this task under manual conditions (64 minutes) compared to the condition where they were supported by the NC system (100 min), $t(13) = 4.31, p = .001$.

In contrast, the overall quality of the surgeries did not differ. This was revealed by almost equal expert ratings for modules which had been milled manually (2.57) and those milled with NC support (2.36), $t(13) = .68, n.s.$

However, the most important difference between conditions emerged with respect to the number of injuries. As expected no injuries of risk structure were found for NC assisted surgeries. In contrast 21% ($n=3$) of the surgeries performed manually resulted in an injury of the sigmoid sinus which in reality would have had severe consequences for the patient. This advantage of NC support was also reflected in expert ratings of possible complications which might have arisen from the surgery. Averaged across all kinds of complication this possibility was rated higher for manual (1.43) than NC assisted conditions (1.20), $t(13) = 2.62, p < .03$.

Workload

Effects of the two experimental conditions on subjective workload assessed by the NASA TLX are shown in figure 1. A significant difference between the two experimental conditions was found for overall workload defined as the average rating across the different dimensions of TLX, $t(13) = 3.83, p < .01$. Participants reported higher workload in the NC condition (13.5 on the scale from 0 to 21) than in the manual condition (9.3). More detailed analyses revealed that this difference was mainly determined by a considerably higher frustration level while working with the support system (15) in contrast to manual milling (6.4), $t(13) = 4.82, p < .001$. All other dimensions of the NASA TLX pointed in the same direction but differences did not become significant if adjustments of alpha-level were taken into account. Observations during the different experimental trials suggested that the different frustration levels were mainly related to a high number of stopping events in the NC condition that were not related to the protection mechanism but to other causes (e.g. loss of line-of-sight). In this case participants were interrupted without being able to easily cross-check the reason for the interruption which would have meant to look up from the microscope and to analyse the navigation display.

This finding was supported by the performance effects found in the secondary task which also point to a higher workload in the NC condition. These effects are shown in figure 2. Compared to the single-task response times (mean: 629 ms), participants needed almost as twice as long (1095 ms) for responding to the acoustical probe during milling. However, this effect was more pronounced in the NC condition. A 2(Condition) x 2(Part 1, i.e. before SAGAT vs.

Part 2, i.e. after SAGAT) ANOVA revealed that participants needed significant more time to respond to the acoustic probe in the NC condition (mean: 1151 ms) than when milling manually (1039 ms), $F(1,13) = 7.94, p < .02$.

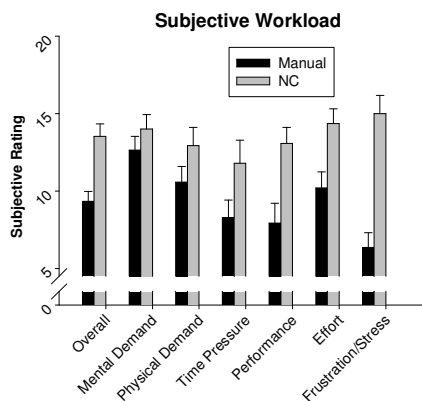


Figure 1: Effects of experimental conditions on subjective workload.

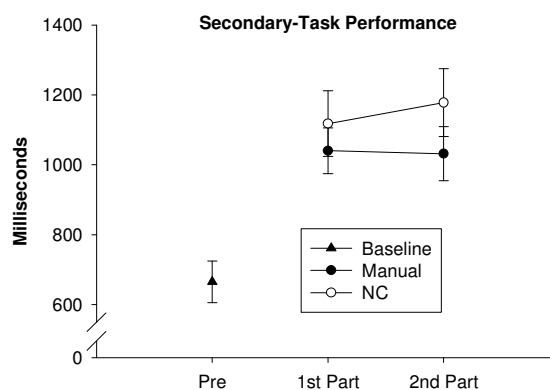


Figure 2: Effects of experimental conditions on secondary task performance.

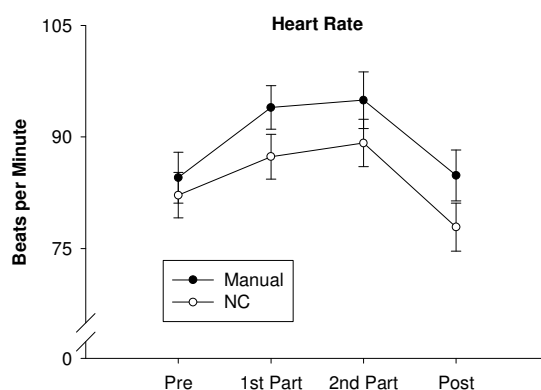


Figure 3: Effects of experimental conditions on mean heart rate.

Despite high level of frustration and longer reaction times in secondary task all physiological data pointed to a lower physiological stress level while milling with NC compared to the manual condition. However, only effects for mean heart became significant. These effects are displayed in figure 3 and were analysed by a 2(Condition) x 4(Time-of-Measurement: pre-baseline, part1, part 2, post-baseline)

ANOVA. This analysis revealed a main effect for Condition, $F(1, 12) = 7.29$, $p < .02$, a main effect for Time-of-Measurement, $F(3, 36) = 45.90$, $p < .001$, and a marginal significant interaction effect, $F(3, 36) = 2.80$, $p < .06$.

As becomes evident from figure 3, heart rate showed a stronger increase in the manual condition compared to NC, and this difference also persisted in the post-baseline measurement. In addition, a significant interaction effect also was found for respiration rate, $F(3, 36) = 5.63$, $p < .01$, reflecting a steeper increase of respiration rate in the first part of the intervention in the manual compared to the NC condition. All other physiological measures but blood pressure only indicated main effects of Time-of-Measurement (all $F < .05$).

Situation awareness

In both conditions participants were able to estimate the relative distances from the milling head to the risk structures at the moment of interruption. Rank correlations between subjective estimates and real differences measured based on screenshots were $r = .47$ (manual) and $r = .51$ (NC). Similarly no significant effects emerged for awareness of already accomplished task goals. However, a marginal significant difference was found with respect to the assessment of anatomical characteristics. Participants in the manual condition were better able to assess specifics of the anatomy (1.64 of 3 possible right answers) than participants working with NC support (1.0), $F(1, 13) = 3.24$, $p < .10$, $\eta^2 = .20$. This effect was most obvious for the assessment of location of the *dura mater* (5 participants gave right answers in manual condition and only 1 in NC condition) and of the *sigmoid sinus* (12 vs. 6 participants), although the location of the latter was directly relevant for the intervention. This behaviour is self-evident, because the participants could not reach *dura mater* and *sigmoid sinus* in the NC condition and thus assess the location. There was no considerable difference in assessment of level of pneumatization (6 vs. 7).

DISCUSSION

The present study provides first insights into human performance consequences of a new computer-based support technology for surgeons. Three different issues were considered: surgical performance and patient safety, different aspects of workload, and situation awareness.

With respect to effects on surgeons' performance a mixed pattern of results emerged. On the one hand, the availability of NC support contributed to an increase of patient safety. In the condition with NC support, even comparatively inexperienced surgeons were able to carry out the complex surgical intervention without any intra-operative injuries of important risk structures while reaching equal milling quality. This does not only support the effectiveness of NC support for protection of risk structures but also suggests that this kind of support can compensate for differences in surgical experience of surgeons. On the other hand, the efficiency, e.g. time needed for the intervention, increased with NC support. This effect resembles earlier results found for pointer-based tools

(Manzey et al., 2009). However, in contrast to the earlier result which seemed to be related to the need of handling additional equipment (pointer), the time-costs for NC were mainly due to events where the trephine stopped due to technical reasons (e.g. loss of line-of-sight).

This effect of interruption of normal workflow due to technical problems might also be responsible for the observed increase of subjective workload and the decrease of secondary-task performance in the NC condition. Participants obviously got distracted by the repeated interruptions of workflow and, thus, were less able to respond to additional tasks. However, a clear beneficial effect of NC support emerged with respect to a lower physiological stress level. This might be explained by the fact that the NC system effectively prevented the participants to come too close to risk structures which lowered the perceived and actual risk of injuring important risk structures which would have had serious consequences for the health of the patient.

Only some marginal effects of NC support were found for situation awareness (SA). Contrary to expectations, SA of surgeons working with NC support was not impaired except some deficiencies with respect to the assessment of specific anatomical characteristics. However, the latter effect was very weak and might be more related to the little experience of the surgeons than on the NC technology.

In summary, the results of the present study suggest that NC technology has a large potential to increase patient safety and to release the physiological stress level of surgeons. Negative performance consequences found with respect to perceived workload and capture of attentional resources do not seem to be related to the general principle of NC support but limitations of the given technological implementation. Improvements and may be a remodelling of the control function of the NC technology might reduce these disadvantages in future.

REFERENCES

- Endsley, M. (2000). Direct measurement of SA: validity and use of SAGAT. In M.R. Endsley & D.J. Garland (eds.), *Situation awareness: analysis and measurement*. Mahwah: Lawrence Erlbaum.
- Hart, S. G., & Staveland, L. E. (1988). Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp.139-183). Amsterdam, The Netherlands: Elsevier.
- Koele, W., Stammberger, H., Lackner, A. & Reittner, P. (2002). Image guided surgery of paranasal sinuses and anterior skull base-five years experience with the InstaTrak-System. *Rhinology*, 40, 1-9.
- Manzey, D., Strauss, G., Trantakis, C., Lueth, T., Roettger, S., Bahner-Heyne, J.E., Dietz, A. & Meixensberger, J. (2009). Automation in surgery: A systematic approach. *Surgical Technology International*, XVIII.
- Manzey, D., Roettger, S., Bahner-Heyne, J.E., Schulze-Kissing, D., Dietz, A., Meixensberger, J. & Strauss, G. (2009). Image-guided navigation: the surgeon's perspective on performance consequences and human factors issues. *International Journal of Medical Robotics and Computer Assisted Surgery*, 5, 297-308.
- Parasuraman, R., Sheridan, T. & Wickens, C.D. (2000). A model of types and levels of human interaction with automation. *IEEE Transactions of Systems, Man, and Cybernetics. Part A – Systems and Humans*, 30, 286-297.
- Strauss, G., Koulechov, K., Hofer, M., Dittrich, E., Grunert, R., Moeckel, H., Muller, E., Korb, W., Trantakis, C., Schulz, T., Meixensberger, J., Dietz, A., Lueth, T. (2007). The navigation-controlled drill in temporal bone surgery: a feasibility study. *Laryngoscope*, 117, 434-41.