An observation support system with an adaptive ontology-driven user interface for the modeling of complex behaviors during surgical interventions

T. Neumuth, B. Kaschek, D. Neumuth, and M. Ceschia
University of Leipzig, Leipzig, Germany

J. Meixensberger and G. Strauss
University of Leipzig, Leipzig, Germany
and University Hospital Leipzig, Leipzig, Germany

and

O. Burgert
University of Leipzig, Leipzig, Germany

The field of surgical interventions emphasizes knowledge and experience; explicit and detailed models of surgical processes are hard to obtain by observation or measurement. However, in medical engineering and related developments, such models are highly valuable. Surgical process modeling deals with the generation of complex process descriptions by observation. This places high demands on the observers, who have to use a sizable terminology to denominate surgical actions, instruments, and patient anatomies, and to describe processes unambiguously. Here, we present a novel method, employing an ontology-based user interface that adapts to the actual situation and describe the principles of the system. A validation study showed that this method enables observers with little recording experience to reach a recording accuracy of >90%. Furthermore, this method can be used for live and video observation. We conclude that the method of ontology-supported recording for complex behaviors can be advantageously employed when surgical processes are modeled.

Surgery requires considerable experience and skill. The surgical process involves highly complex procedures and depends on a multitude of factors that entail the surgeon's awareness and attention to patient-specific abnormalities in anatomy and pathology or technical resources. With respect to the multifaceted and highly variable processes of surgical interventions, a modeling and behavior observation strategy is necessary to account for this complexity. Until today, a concise form of report that is able to reproduce surgical process evolution in a detailed and accurate way has been missing. A comprehensive observation and modeling tool that could record the progress of processes such as surgical interventions would be a novel and insightful approach to this challenge.

The availability of expressive process models supports a multitude of other application areas in medical engineering, aside from medical training—most notably the performance of requirements analyses (Neumuth, Trantakis, et al., 2009) or the introduction and evaluation of new therapeutic standards and surgical assist systems (Strauss et al., 2006). Evolutionary models of surgical processes, surgical process models (SPMs), can further the under-

standing, reproduction, analysis, training, and teaching of surgery. However, to be used in the aforementioned application areas, the recordings and observations must be accurate and comprehensive.

In the literature, various works have focused on the modeling of process sequences in surgery. Some research groups have used interviews as the main basis for their investigations (Jannin, Raimbault, Morandi, Riffaud, & Gibaud, 2003; Raimbault, Morandi, & Jannin, 2005), whereas others have proposed the use of dedicated measuring systems, which record partial information about surgical process steps (Ahmadi et al., 2006; James, Vieira, Lo, Darzi, & Yang, 2007; Padoy et al., 2007). However, the most frequently used techniques are based on observers and range from data acquisition without software support (den Boer et al., 1999; Malik, White, & Macewen, 2003; Mehta, Haluck, Frecker, & Snyder, 2002; Strauss et al., 2006) to observer-based methods that are supported by software systems (Neumuth, Jannin, Strauss, Meixensberger, & Burgert, 2009). Moreover, several software solutions (Castellano, Perea, Alday, & Mendo, 2008; Hänninen & Pastell, 2009; MacLin & MacLin, 2005) and

combined hard- and software solutions (Held & Manser, 2005; Sarkar et al., 2006) have been proposed in the context of behavior analysis.

The procedure of modeling surgical processes by observation, especially with the objective of performance assessment in the context of surgical training (Leong et al., 2007; Megali, Signigaglia, Tonet, & Dario, 2006; Richards, Rosen, Hannaford, Pellegrini, & Sinanan, 2000; Rosen, Brown, Chang, & Hannaford, 2006) or of comparing strategies for surgical treatment (den Boer et al., 1999; Strauss et al., 2006), generally focuses on two measurement strategies: high-resolution, sensor-based measurement of the performance of a limited number of surgical actions—such as, for instance, instrument movement trajectories while placing knots—or low-resolution and simpler observer-based measurements for surgical interventions or interventional phases (e.g., Archer & Macario, 2006; Schuster, Wicha, Fiege, & Goetz, 2007), without reference to specific surgical process steps. In previous work, we proposed an approach that allows for a medium level of granularity to be used for the decomposition of surgical process steps into categories, as described in the following section, in order to accommodate the complexity and diversity of information and the high variability of surgery. An observation support software, known as the Surgical Workflow Editor (Neumuth, Czygan, Strauss, Meixensberger, & Burgert, 2009; Neumuth, Durstewitz, et al., 2006; Neumuth, Jannin, et al., 2009; Neumuth, Strauss, Meixensberger, Lemke, & Burgert, 2006), was developed for such observation on a medium level of granularity.

Due to the diverse information that can be acquired with the observation approach for surgical processes on a medium level of granularity, high demands are put on observers, because, apart from being under continuous time pressure during the observation, they also need to deal with extensive surgical and anatomical terminologies. For this reason, we introduce the usage of ontologies as knowledge bases to support the observer by means of an adaptive user interface, a situation-dependent edition of the observation software interface. The use of a knowledge-based system is necessary to overcome initially contradictory objectives posed by the new observation strategy: On the one hand, due to the high variability of surgical processes, the recording system needs to be able to deal with a large number of surgical terms to achieve expressiveness, and, on the other hand, this information needs to be declarative to abstract higher resolution information and to allow for usability by clinical and nontechnical users.

To our knowledge, there is currently no method available that deals with the support of the observer using knowledge-based software, although adaptive software interfaces are in use in several applications in computer science (Berrais, 1997; Kuehme, 1993; Love, Jones, Tomlinson, & Howe, 2008). With the help of such interfaces, the amount of terms available in observation situations can be considerably decreased and, thus, reduce the observers' workload. An implementation of the proposed method could also be employed to great advantage in other fields of science that rely on data gathering as the basis for observation or monitoring, such as behavioral or educational science.

This work presents the working principle of the observation support system with the adaptive user interface, shows its technical implementation, and presents the results of a validation study that demonstrates the accuracy of the system. Within the validation study, it is shown that the system is applicable to live and video-based observation. Furthermore, the study introduces several measurements for assessing granularity, content accuracy, and temporal accuracy of observed surgical process steps, as well as observer workload. The study is based on simulated interventions from the field of otorhinolaryngology; an application to other surgical disciplines or for use in cases of behavior research is possible but is not discussed explicitly.

METHOD

### Development of Ontology and Materials Used for Testing

Some specific terminology will be used and explained in this section. Furthermore, the software that has been used will be presented, and its functionality explained.

Surgical process models. Following the Workflow Management Coalition's (1999) definition of a business process, we define a surgical process (SP) as *a set of one or more linked procedures or activities whose instances (are intended to) collectively realize surgical objectives within the context of an organizational structure defining functions, roles, and relationships*. An SP is transformed into an SPM, a model representing the surgical intervention. The SPM as the core concept of our approach is derived via an observation protocol and represents partial aspects of the original SP in a formal or semiformal way.

The surgeon's work is rendered as temporally extended process steps in the SPM, called *activities*, which consist of various perspectives (Jablonski & Bussler, 1996). All perspectives conjointly form an activity. Each of the activities describes a different point of view on a surgical process step: The organizational perspective describes who performs a process step, such as "surgeon" or "assistant"; the functional perspective describes what is done in a process step, such as "cutting" or "suturing"; the operational perspective indicates the technical resources that are used to perform a process step, such as "scalpel" or "needle"; the spatial perspective describes at which location at the patient's body the process step is performed, such as at the "sinus maxillaris"; and the behavioral perspective indicates at which point of time a process step takes place; this perspective is represented by time stamps.

Observation support software system with an adaptive user interface. The established way of composing SPMs is to record them with the help of specially trained observers. The observer relates terms concerning the several perspectives to each other—for example, actions are assigned to single members of the operating room staff—in order to create a description of a situation. Herein, he is supported by the observation support software, the Surgical Workflow Editor, which generates the SPM as an observation protocol. The Surgical Workflow Editor is a JAVA-based application used for the recording and analysis of SPMs. It stores such information as is acquired by the observer, with the help of an interface that displays all possible entities needed to record surgical interventions, such as surgical instruments and activities.

However, there are some general difficulties in recording SPMs on this medium level of granularity. First, a single surgical intervention can consist of up to 300 single-process steps, each of which has different perspectives. Second, the observers need to choose the proper terms for the different perspectives out of large repositories: The designated terminology for the operational perspective might comprise about 50 different surgical instruments; the terminology for the functional perspective might include about 30 different actions; and about 20 different anatomical structures might be used for the spatial perspective. Third, a preliminary definition of the

A

B
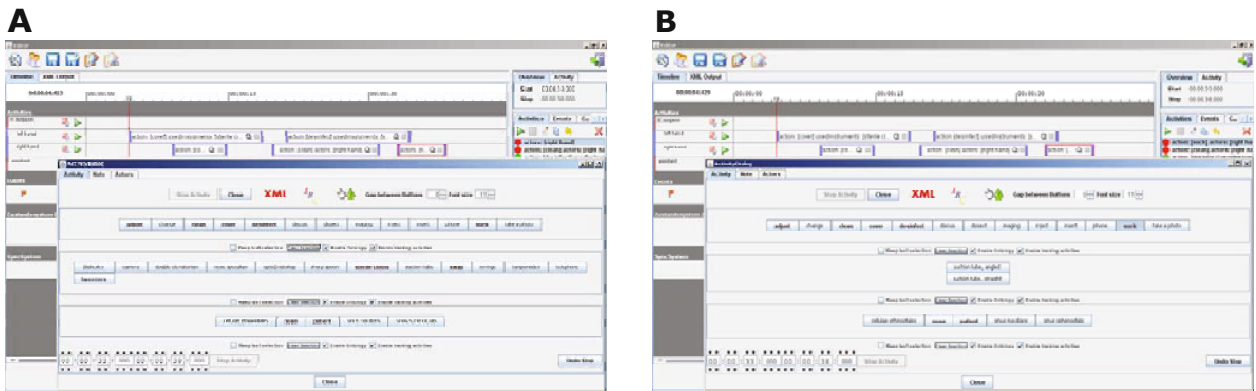
Figure 1. Functionality of the ontology-based adaptive user interface: After the action "suck" is selected (A), the terminology list for surgical instruments is restricted automatically to "suction tubes" (B). In contrast to the static interface of the conventional system, the user has to choose the correct instrument from a smaller variety of items.

different observational codes, as peculiar to the realm of behavioral research, is not possible here, due to the sheer number of potential combinations of terms from each of the perspectives. And, lastly, the display size of the computer screen limits the choice of representable items from the terminology.

One possible approach to overcoming these problems is the implementation of an ontological knowledge base support. An ontology is an "explicit specification of conceptualization" (Gruber, 1993), a formal representation of concepts and their relations. The perspectives used to describe process steps in our approach, instruments, activities, and anatomical structures were defined as ontological concepts by domain experts, and logical conjunctions between different elements were described as relations that link terms to one another.

Various tools are available to design these logical constructions. We used Protégé (Stanford University, 2009), which formalizes the concepts and the semantic relationships between the concepts and allows for an implementation of enhancing applications, such as logical reasoners that test the relations for formal correctness (Racer Systems, 2009). In the proposed system, the generated Protégé ontology in OWL format (Web Ontology Language) is saved into the ontology server and is loaded when the observation support software is started.
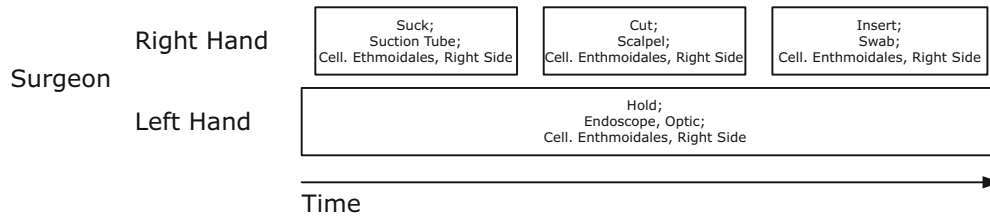
This knowledge base support was designed to facilitate observations for the recorder and to allow the choosing of the appropriate term that best describes the most recent process step without browsing extensive terminology lists. Depending on the entity selected from one perspective, only terms from the other perspectives that can be sensibly combined with the preselected perspective are shown. For example, if "cutting" were chosen as the functional perspective for a surgical activity, the list of surgical instruments would be restricted to those that are "able to cut"—for instance, "scalpel" and "scissors." An illustration is provided in Figure 1.

More precisely, our technical solution consists of three main components (see Figure 2): the ontology server that contains the knowledge base, the adaptive user interface, and the editor engine. The knowledge base contains the concepts and the relations between the concepts that are necessary to record a surgical intervention. The adaptive user interface is the input mask for the instantiation of items by the user and represents the contents of the perspectives. It adapts automatically to the current situation, according to the actual user input and the knowledge base response. The editor engine administers and delegates the central business logic, such as the communication management with the ontology server, for example.

Surgical application case and simulation scripts. System validation used functional endoscopic sinus surgeries (FESS) as an
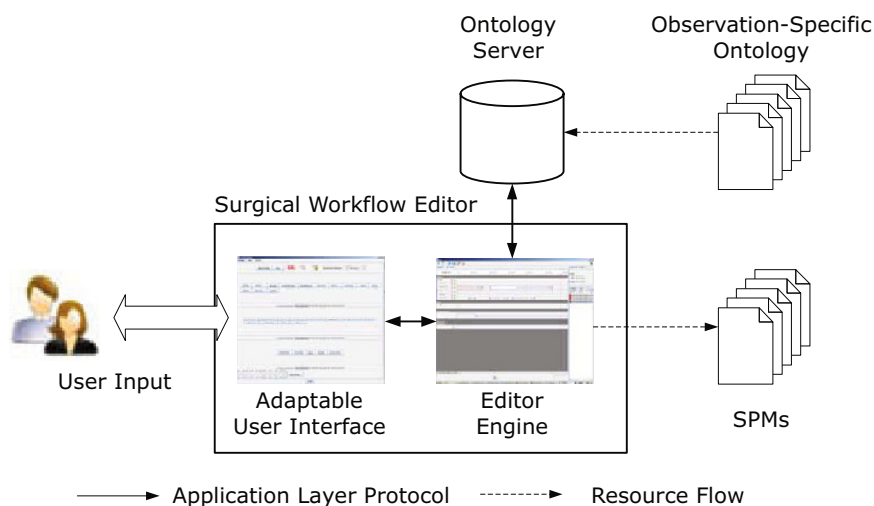
Figure 2. Software component infrastructure and data flow for the adaptive knowledge-based observation support system.

	Right Hand	Suck; Suction Tube; Cell. Ethmoidales, Right Side	Cut; Scalpel; Cell. Enthmoidales, Right Side	Insert; Swab; Cell. Enthmoidales, Right Side

Surgeon

Left Hand — Hold; Endoscope, Optic; Cell. Enthmoidales, Right Side

Time →

Figure 3. Cutout procedure of simulation script for the functional endoscopic sinus surgeries simulation with detailed instruction for the actors.

example. In Germany, it is one of the most frequently performed interventions in surgical otorhinolaryngology: About 50,000 such procedures are performed annually (German Federal Statistical Office, 2008). This minimally invasive, endoscopic intervention involves examining nasal cavities, such as the ethmoid sinuses, and removing diseased or obstructive tissue or growths, such as nasal polyps. The goal of the FESS is to purposefully remediate diseased areas and restore natural drainage and ventilation paths. During the intervention, the surgeon generally handles the endoscope with one hand, while the other hand performs surgical activities, such as the removing of tissue with forceps or exhausting liquids.

The basic purpose of the validation study was to design and act out FESS interventions according to simulation scripts, as is shown in Figure 3. The simulation scripts were developed by ICCAS Institute in close cooperation with surgeons from the Department for Otorhinolaryngology at the University Hospital Leipzig. Three simulation scripts were developed, each the variant of a different typical type of FESS surgery. These simulation scripts contained detailed patterns for process steps of FESS interventions for two actors ("surgeon" and "assistant"), 16 surgical instruments, 13 functional tasks, such as "cut" and "clean," and seven anatomical and pathological structures. Figure 3 shows examples of devised process steps. Each simulation script had a total duration of about 20 min and contained 60–90 single-process steps. As preparation for the validation of our system, the three simulation scripts were read aloud and, thus, recorded as audio file instructions.

Participants

For the present study, 6 observers, referred to as *novices*, were recruited to record the SPMs. None of the observers had any previous experience with recording or describing surgical interventions, nor had they used the software before. The 6 novices were then trained to collect data for SPMs in an introductory workshop. Additionally, they were introduced to the topic of FESS interventions, the intervention type's typical progression, its surgical goal, the instruments used, and the anatomical and pathological structures involved. Furthermore, the introductory workshop was used to present the predefined ontology and its coherent use during the observation to the participants.

Out of the 6 novices, 3 had a medical background, studying human medicine, whereas the other 3 had an engineering or computer science background. This distinction was made to test whether observers without medical background can achieve the same quality of results as the medical students, since our experience from the last years has shown that nonmedical students have a less tight curriculum and, therefore, more time for extracurricular activities.

Study Design

The presented recording system has been validated in a complex study setup. On the basis of simulated surgical procedures, observers were asked to generate SPMs as precisely as possible, using the observation software. Subsequently, the accuracy of the observation protocols was compared with the simulation scripts (see Figure 4).

Preparation of data acquisition. The simulation scripts were performed by actors that received instructions about the process steps from an audio file by means of mp3 players and earphones.

The simulation was performed on anatomically correct and detailed paranasal models, as shown in Figure 5.

For the video observations, each of the three simulation scripts was performed once, while being recorded by multiple video cameras. The resulting videos were then cut short and rendered into presentations. Two observation experts having significant experience with the Surgical Workflow Editor generated the SPMs for the filmed simulations and converted the protocols to XML format. These observations were compared with the simulation scripts as references for the validation of the simulation itself.

Data acquisition. Live observation was performed with the help of live-acted simulations, which were journalized by the novices. In addition, these simulations were documented on video. After the recording, the two experts compiled their observations from the video documentation of the simulations as references. These were again compared with the simulation scripts for the validation of the simulation.

Data acquisition was performed in the demonstrator OR at the ICCAS of Leipzig University. The overall study took 4 days within a time frame of 2 weeks. Following an introductory workshop, all the observers had to complete live and video observations with tablet PCs thrice for each of the simulation scripts (see Figure 6). The succession of live and video observations alternated, and the sequence of the simulations was randomized. To summarize, it can be said that each observer recorded all three simulations thrice, resulting in 9 protocols. This holds true for the live and the video recordings, accounting for a total number of 18 protocols. Thus, considering all 6 observers, an overall amount of 108 protocols was achieved.

Postprocessing and analysis. After data collection, the SPMs of the simulation scripts, the observations by the experts, and the observations by the novices were transferred to a PostgreSQL 8.3 database (PostgreSQL Global Development Group, 2009). The appraisal followed the three-stage concept of the study. In preparation for validation, the observations by the experts and the observations by the novices were compared to validate the accuracy of the observations. Furthermore, each of the observations by the experts was compared with its corresponding simulation script to validate the simulation accuracy. This was accomplished with the help of a special software tool, which represented each recorded activity in relation to its respective reference. The experts could then decide manually whether or not the novices had recorded the right action, instrument, and anatomical structure. The observations by the experts were employed to retain differences between simulations and the observations by the novices, which occurred when actors made mistakes during the simulated FESS interventions.

The data acquisition method (live or video recording), the background of the observers (medical or engineering), and the number of the simulation script were regarded as independent variables. As for the validation of the accuracy of the knowledge-based observer support system, five different dependent variables were analyzed: granularity, content accuracy, temporal accuracy, completion time, and workload for the observers.

The goal of measuring the granularity of the recorded process steps was to determine the ratio of structurally correctly recorded activities in the observations by the novices in reference to the respective observations by the experts. A correct granularity would be a 1:1 re-

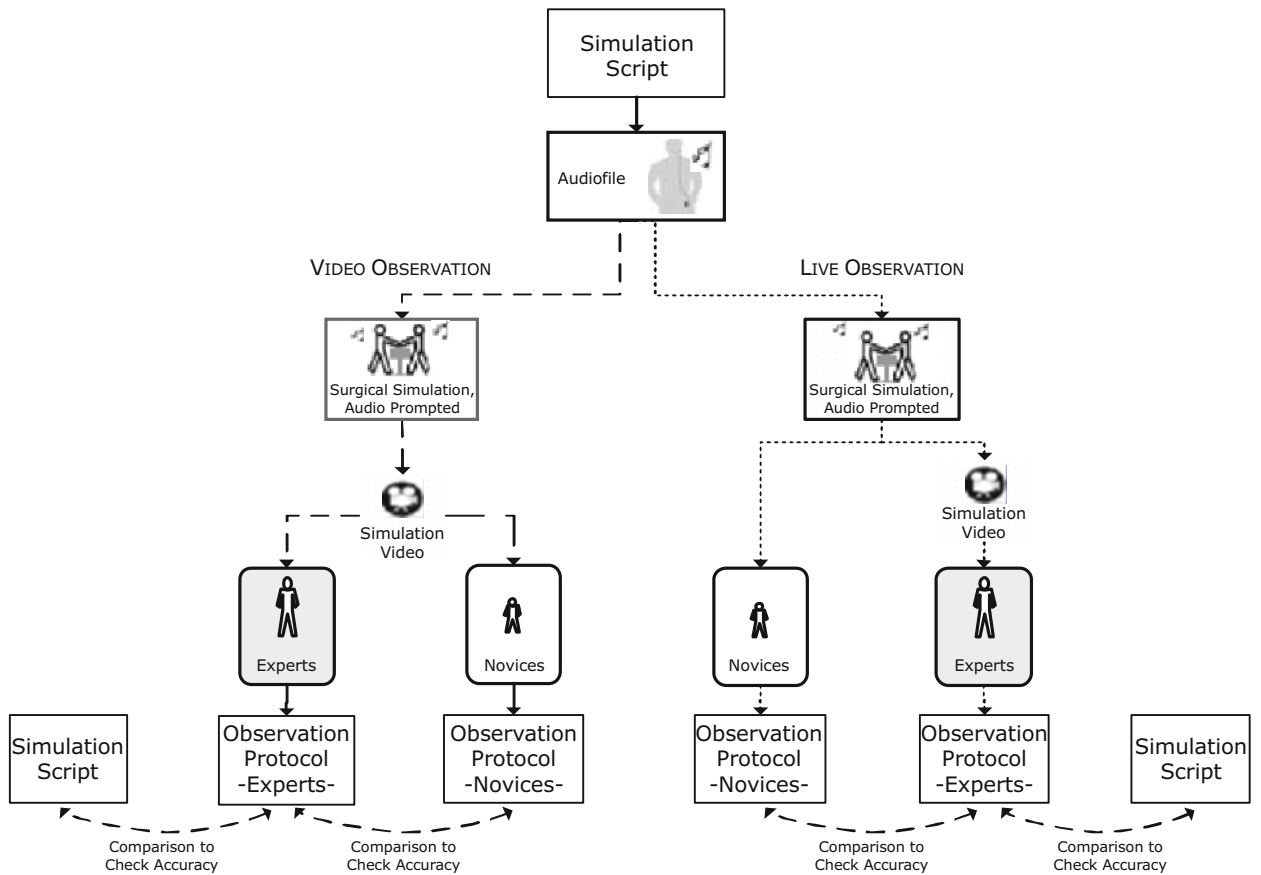**Figure 4. Overview of the system validation study showing the data flow from the design of the simulation scripts to the protocols.**
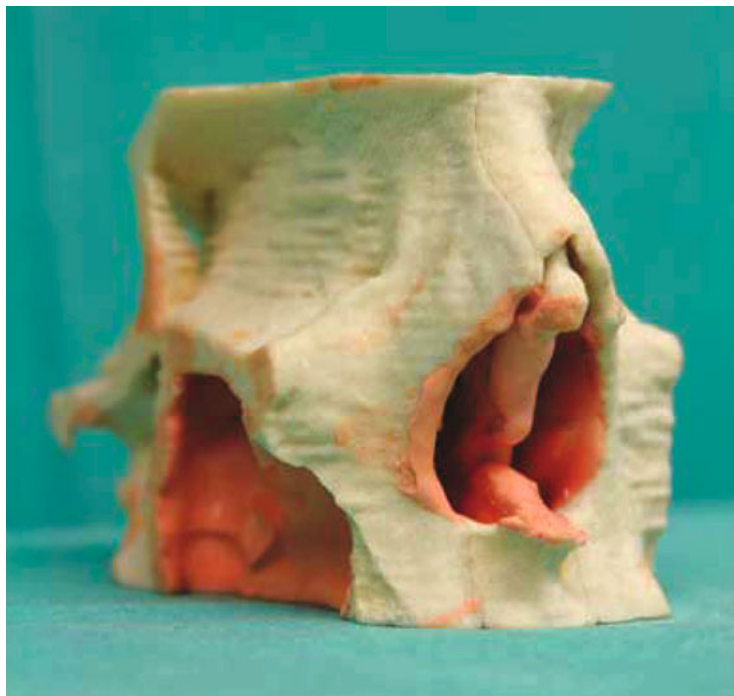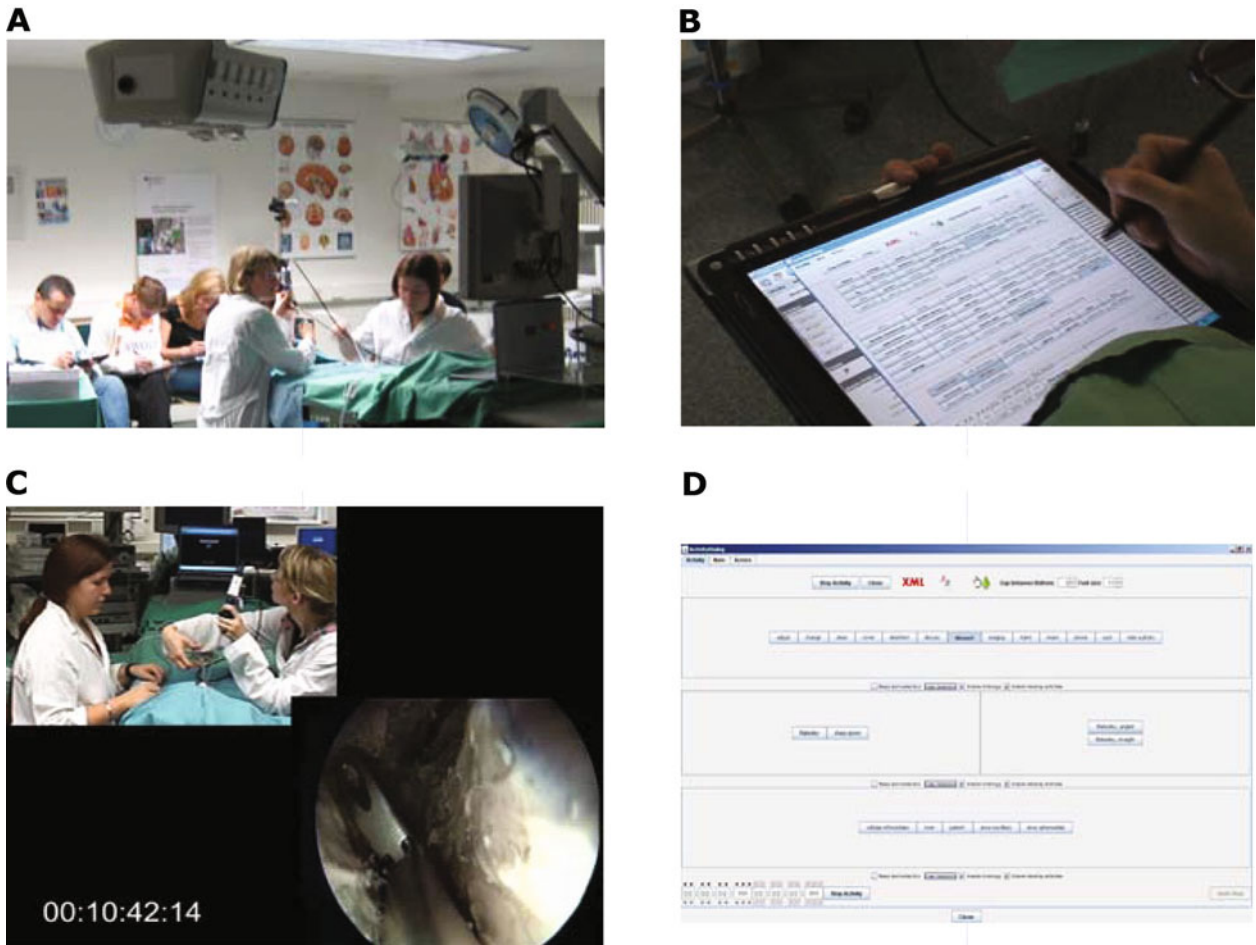

Figure 5. 3-D print model of the nasal cavities of a real patient as used in the system validation study.

Figure 6. Data acquisition example. Novices watching the simulation scripts performed by the actors (A) and operating a tablet PC (B) to record the surgical process. (C) The surgical process step "dissection of material with a Blakesley" (a specific type of surgical pliers) is seen as performed by the actors and on the endoscope screen. (D) The observer selects "Blakesley" or "dissect" in the observation support software, and the interface adapts according to the relation between "Blakesley" and "dissect."

lationship between the recorded activity and the respective activity in the reference. Incorrect activities were regarded as additional observations that did not appear in the reference (0:1 relation) or missing observations that were missed in the observations by the novices but appeared in the observation by the experts as reference (1:0 relation). Other possibilities were the logging of single activities as multiple activities (increased granularity; $1:m$, $m > 1$), the subsumption of different activities into a single activity (decreased granularity; $n:1$, $n > 1$), or mixed granularity of activities ($m:n$; $m \neq n$, $m, n > 1$).

The content accuracy was determined by a comparison of the content of perspectives in the observation protocol with the perspectives in the particular references. A total correlation was appraised as 1, and a deviation as 0. Subsequently, the percentage of correctly observed activities for the respective perspective was determined.

The temporal accuracy identified the absolute value of temporal deviation between the duration of activities in the observation and reference. The measurement of the completion time is expressed as a ratio between the duration of the observation and the duration of the simulation. The observation duration measured the expenditure of time needed to complete the whole protocol by indicating the temporal margin between the start of the recording and the final release of the workflow protocol by the observer.

In addition to examining the outcome of the observation, we have tested the usability of the observation support system with the adaptive user interface. For this goal, the NASA task load index (TLX;

Cao, Chintamani, Pandya, & Ellis, 2009; Hart & Staveland, 1988) was employed, which was specifically intended for operators of human–machine systems and which gathered subjective information about physical workload. The TLX includes six subscales: Mental Demands, Physical Demands, Temporal Demands, Performance, Effort, and Frustration. The novices were asked to fill out a questionnaire after each observation.

The statistical analysis was conducted with the help of a GLM (Generalized Linear Model) for the independent variables of live or video observation, medical or engineering observers, and simulation script. All statistical tests were conducted with SPSS 15 (SPSS Inc., Chicago) at a significance level of $\alpha = .05$.

RESULTS

In the context of the validation study, the novices had to record SPMs in live and video observations according to the methods described in the previous section. Data acquisition was performed using the adaptive user interface of the Surgical Workflow Editor. The results are presented in Table 1.

The observers reached a mean correct granularity of 92.8% ± 7.3% (mean ± standard deviation). With regard

Table 1
Study Results for Granularity, Content Accuracy, and Observer Workload ($df = 106$)

| | Data Acquisition Strategy | | | | | | Observer Population | | | | | | Gold Standard | | | | | | | |
| | Live | | Video | | Significance | | Medical | | Engineering | | Significance | | #1 | | #2 | | #3 | | Significance | |
	M	SD	M	SD	F Value	p	M	SD	M	SD	F Value	p	M	SD	M	SD	M	SD	F Value	p
Granularity																				
Additional observation	0.5	0.8	0.7	1.0	3.1	.08	1.1	1.1	0.1	0.3	**54.3**	**<.001**	0.6	1.0	0.8	1.0	0.4	0.8	2.9	.06
Missing observation	3.6	7.3	2.2	3.9	2.1	.15	2.0	2.9	3.8	7.7	3.3	.07	2.3	5.1	0.6	1.1	5.8	7.9	**8.7**	**<.001**
Correct granularity	90.5	10.1	94.5	5.9	**7.6**	**.007**	94.4	5.3	90.6	10.5	6.9	.01	92.9	9.0	95.6	4.8	89.0	9.6	**6.5**	**.002**
Decreased granularity	0.6	1.3	0.8	1.9	0.3	.59	1.1	2.1	0.3	0.8	**8.6**	**.004**	1.0	2.0	0.1	0.7	1.0	1.7	3.8	.03
Increased granularity	5.7	6.9	3.1	4.8	4.8	.03	2.6	4.2	6.2	7.0	**10.5**	**.002**	4.6	7.2	3.9	4.9	4.7	5.9	0.2	.84
Content Accuracy																				
Functional perspective	97.9	1.8	97.5	2.1	4.6	.04	97.6	2.2	97.8	1.8	2.4	.12	98.3	1.6	98.1	1.6	96.7	2.4	**27.2**	**<.001**
Organizational perspective	98.1	2.8	98.8	1.6	1.7	.20	98.7	2.2	98.2	2.3	0.1	.73	99.2	1.4	99.5	0.9	96.7	2.9	**6.3**	**.003**
Operational perspective	86.5	6.6	85.6	5.1	0.9	.33	84.3	6.7	87.8	4.2	**12.8**	**.001**	86.5	3.9	89.3	5.8	82.3	5.6	**15.9**	**<.001**
Spatial perspective	75.5	5.6	73.7	5.9	5.0	.03	73.8	4.5	75.5	6.8	3.7	.06	73.6	4.4	80.0	3.7	70.3	4.5	**47.9**	**<.001**
Total content accuracy	92.9	2.9	92.7	2.5	0.2	.67	92.2	2.6	93.5	2.7	6.7	.01	93.2	2.2	94.1	2.3	91.1	2.7	**13.7**	**<.001**
Observation Workload																				
Effort	56.4	18.0	51.2	20.4	2.8	.10	42.1	15.9	65.5	14.9	**58.6**	**<.001**	55.0	19.0	52.0	20.7	54.3	18.7	0.5	.62
Frustration	48.8	18.0	41.1	17.8	5.1	.03	39.6	15.8	50.1	19.0	**9.2**	**.003**	42.6	16.7	44.0	17.9	48.2	20.1	1.0	.36
Mental demand	51.2	18.5	46.3	20.1	3.5	.07	35.2	14.5	62.2	13.5	**95.0**	**<.001**	46.0	19.1	49.6	19.8	49.9	19.8	0.4	.66
Performance	43.7	17.6	38.3	15.9	2.5	.12	40.1	12.0	41.8	20.8	0.2	.65	40.9	17.6	40.6	16.7	41.3	16.8	0.0	.99
Physical demand	48.3	15.7	45.8	16.9	0.6	.45	46.1	16.6	47.9	16.1	0.3	.57	45.8	15.6	47.0	17.6	48.4	16.0	0.2	.80
Temporal demand	56.1	17.5	45.6	18.6	**11.9**	**.001**	41.1	15.2	60.4	17.0	**38.7**	**<.001**	50.1	17.9	51.0	19.2	51.2	19.6	0.1	.94
Total workload	54.8	13.6	49.6	14.4	5.1	.03	44.2	10.6	60.1	12.8	**47.1**	**<.001**	51.4	13.4	51.7	14.7	53.4	14.7	0.3	.74

Note—Granularity and observation accuracy values are given as percentages. Observation workload is a unitless value on a scale between 0 (*no demand*) and 100 (*highest demand*). Significant results with $p < .01$ are printed in boldface.

to granularity, video observation was 8.6% more accurate than live observations. Furthermore, medical students recorded surgical behavior with 8.4% more accurate granularity than did engineering students. In contrast, decreased granularity was more frequent in live observation protocols. No mixed granularity was observed.

The total content accuracy was 92.3% for the medical group, whereas the engineering group achieved 93.5%. Arguably, the main differences originated from the operational perspective. The temporal fidelity of the activities in the observation protocols showed a comprehensive mean error of 2.0 ± 1.6 sec and no statistically significant difference between the groups.

Only slight disparities were observed between the observer populations with regard to completion time, as is shown in Figure 7. However, the difference between the data acquisition strategies live and video observation in this respect was highly significant ($p < .001$).

Interestingly, all workload criteria were rated as more demanding for live observations than for video observations. Nevertheless, the differences were not significant, with the sole exception of temporal demand ($p < .001$). Observers with an engineering background rated the overall workload higher than did medical observers, in all regards.

The results for the different simulation scripts showed small differences between the means. The observation of simulations for one of the scripts showed slightly less accuracy concerning granularity and content accuracy.

Between-subjects effects, such as the combination of data acquisition strategy and observer population, were not significant.

DISCUSSION

In the research field of medical engineering, the employment of observational strategies is highly relevant—for instance, to obtain SPMs as a base for performing requirements analyses for surgical assist systems or the evaluation of newly developed surgical instruments.

The objective of this work was the implementation of a knowledge-base-driven adaptive user interface for the Surgical Workflow Editor observation support software that provides assistance to observers who need to deal with large terminologies. As we have shown, this method constitutes a robust and expressive basis for the observation of highly variable surgical behavior.

As the sample application for the proposed methodology, a validation study for the modeling of surgical interventions was designed and performed. The study results have shown that even inexperienced observers were able to attain good results by means of the new tools presented. The utilization of this method was nearly equivalent for live and video observations. Even lay users and anatomically or medically inexperienced persons attained good results with this method. However, utilizing this methodology in other fields, such as behavioral science, is also very likely and recommendable.

The results of the validation study for the proposed system showed that it is eligible for both video and live observations, because only a few significant differences were found between these data acquisition strategies. A slightly higher correct granularity of activities was achieved in video observations. Differences between medical and engineering observers were significant for several criteria. However, for the accuracy criteria, these differences were mostly less than 5%.

Observers who have to deal with large terminologies can be adequately supported by engineering systems for recording surgical processes. This work presented a methodology for the application of a knowledge-base-driven adaptive user interface for observation support software. This adaptive user interface represents terminology information depending on the current situation, and the system is employed to support the user in modeling surgical interventions.

Due to the refinement of surgical activities into different perspectives, which is required because of the variability of surgical processes, a multitude of possible combinations of terms emerges. In the moments of recording, the observer
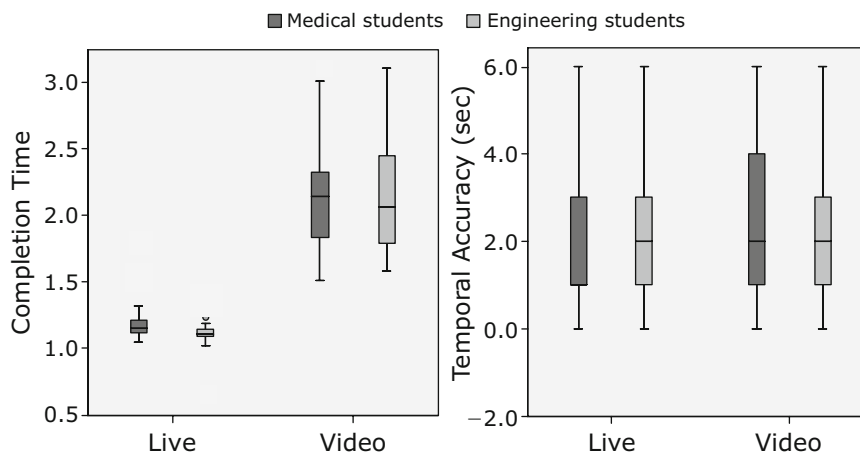

Figure 7. Study results for the completion time (left) and temporal accuracy (right) in seconds. The boxplots provide medians and sample ranges for observations by medical and engineering students in live and video settings.

combines these terms from different perspectives into sensible activities that describe the surgical process steps. For this, the implementation of ontological relations between the terms of the different perspectives, such as surgical instruments and actions that can be performed with them, is reasonable. With the employment of ontological relations, large terminologies can be prerendered for observation tasks. This can be very informative, especially with respect to complex or variable environments, as demonstrated here, using the example of surgical interventions. Thus, the diversity of observable entities is no longer limited by engineering factors, such as the account of a large number of observation categories on a small display.

Our observation approach and the analysis are based on an approach with a medium level of granularity. Granularity for surgical work, which is hard to capture in formal ways, strongly relies on subjective estimation. Due to the nature of our study, we defined the granularity of the process steps described in the simulation scripts as "correct" and instructed our observers accordingly, in the training workshops prior to the study.

None of the novices had any previous experience in observing surgical processes or recording with the presented system. In addition, the observers with an engineering background had neither any surgical background knowledge nor any medical training in advance. Nevertheless, as is shown by the performed validation study, both observer populations reached a correct granularity of about 90% and also a total content accuracy of approximately 90%. This result was achieved despite the observers' lacking previous experience in dealing with the observation software and with the subject matter of the observation itself.

Only the recordings of the spatial perspective showed some peculiarities. Here, neither engineering nor medical observers attained acceptable results. A possible reason for this could be the complex anatomic features of the nasal cavities, which make observations highly demanding, even for medical observers. Another reason for the difficulties in determining the correct anatomical locations could be the ontology itself. Even though ontological relations between the surgical actions and the instruments are relatively explicit, the triangle among the functional, operational, and spatial perspectives has to be rather loosely interpreted, because most combinations of surgical actions and instruments can be applied to anatomical structures.

It can also be presumed from the results that the observer training and the performance of the surgical simulations were appropriate. The duration of the training workshop was mostly adequate, as attested by the average accuracy of about 97% for each perspective, with the sole exception of anatomical structures, as previously mentioned. Seemingly, more experience in this field is advisable. The simulation accuracy of the surgical simulations was validated by comparing the observations by the experts and the simulation scripts. Here, it became apparent that the surgical simulation was accurate, achieving 99.9% ± 0% in granularity and 99.5% ± 1.6% in content accuracy. The statistical results for the simulation scripts showed that all three scenarios were commensurable. All the results for the metrics had the same magnitude.

The analysis of the NASA TLX index has shown that the observation support system with the adaptive user interface is an appropriate means for both data acquisition strategies. As was expected, live observations placed higher workload demands on the observers than did video observations. This disadvantage of live observations might be compensated for by the significantly lower completion time for the live observations, which decreases study costs. Nevertheless, it was also shown that the observational workload was significantly higher for nonmedical observers. However, we attribute these results to the fact that the engineering students had to cope with two new challenges, the observation support software and the medical background they needed to acquire, whereas the medical students had only the software as a new challenge, whereas the medical background was known to them beforehand.

The method for the generation of SPMs presented in this article provides a well-founded basis for the observation of surgical behavior, as we have shown with the help of a clinical example. The validation study was conducted on the basis of an intervention in otorhinolaryngological surgery. Similar accuracy would be expected if the results of this study were generalized to other types of surgical interventions, which should be experimentally substantiated.

In addition, an adaptation of the Surgical Workflow Editor software for nonsurgical application fields that require structured observation of behavior is conceivable. Due to the configurability of the software and the possibility of adjusting the ontology, the implementation of this tool can be extended to include a wide range of possible future study fields from diverse areas, such as sociobiology and psychology. Sociobiology—or, more specifically, ethology—deals with ethograms, catalogues of discrete behavior, which could be described with explicit reference to their purpose with the help of this tool. As for psychology, for instance, usages in the fields of behavioral psychology and educational research can be conceived of, as well as industrial or organizational psychology. More specifically, applied behavioral analysis, organizational learning, problem solving, the development of educational technology, and scientific management could benefit from the implementation of the ontology and the software presented here.

REFERENCES

Ahmadi, S., Sielhorst, T., Stauder, R., Horn, M., Feussner, H., & Navab, N. (2006). Recovery of surgical workflow without explicit models. In R. Larsen, M. Nielsen, & J. Sporring (Eds.), *Proceedings of Medical Image Computing and Computer-Assisted Intervention— MICCAI 2006: Part I. Lecture notes in computer science* (Vol. 4190, pp. 420-428). Berlin: Springer.

Archer, T., & Macario, A. (2006). The drive for operating room efficiency will increase quality of patient care. *Current Opinion in Anaesthesiology*, **19**, 171-176.

Berrais, A. (1997). Knowledge-based expert systems: User interface implications. *Advances in Engineering Software*, **28**, 31-41. doi:10.1016/S0965-9978(96)00030-0

Cao, A., Chintamani, K., Pandya, A., & Ellis, R. (2009). NASA TLX: Software for assessing subjective mental workload. *Behavior Research Methods*, **41**, 113-117. doi:10.3758/BRM.41.1.113

Castellano, J., Perea, A., Alday, L., & Mendo, A. (2008). The measuring and observation tool in sports. *Behavior Research Methods*, **40**, 898-905.

den Boer, K., Straatsburg, I., Schellinger, A., de Wit, L., Dankelman, J., & Gouma, D. (1999). Quantitative analysis of the functionality and efficiency of three surgical dissection techniques: A time-motion analysis. *Journal of Laparoendoscopic & Advanced Surgical Techniques*, **9**, 389-395.

German Federal Statistical Office (Ed.) (2008). *International Classification of Diseases 10: National health statistics for 2007*.

Gruber, T. (1993). A translation approach to portable ontologies. *Knowledge Acquisition*, **5**, 199-220.

Hänninen, L., & Pastell, M. (2009). CowLog: Open-source software for coding behaviors from digital video. *Behavior Research Methods*, **41**, 472-476. doi:10.3758/BRM.41.2.472

Hart, S., & Staveland, L. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human Mental Workload*, **1**, 139-183.

Held, J., & Manser, T. (2005). A PDA-based system for online recording and analysis of concurrent events in complex behavioral processes. *Behavior Research Methods*, **37**, 155-164.

Jablonski, S., & Bussler, C. (1996). *Workflow-management: Modelling concepts, architecture and implementation*. London: ITP New Media.

James, A., Vieira, D., Lo, B., Darzi, A., & Yang, G. (2007). Eye-gaze driven surgical workflow segmentation. In R. Larsen, M. Nielsen, & J. Sporring (Eds.), *Proceedings of Medical Image Computing and Computer-Assisted Intervention—MICCAI 2007: Part II. Lecture notes in computer science* (Vol. 4792, pp. 110-117). Berlin: Springer.

Jannin, P., Raimbault, M., Morandi, X., Riffaud, L., & Gibaud, B. (2003). Model of surgical procedures for multimodal image-guided neurosurgery. *Computer Aided Surgery*, **8**, 98-106.

Kuehme, T. (1993). *Adaptive action prompting: A complementary aid to support task-oriented interaction in explorative user interfaces* (Graphics, Visualization and Usability Center [GVU Center] Tech. Rep. GIT-GVU-93-19). Retrieved September 3, 2009, from http://smartech.gatech.edu/handle/1853/3629.

Leong, J., Nicolaou, M., Atallah, L., Mylonas, G., Darzi, A., & Yang, G. (2007). HMM assessment of quality of movement trajectory in laparoscopic surgery. *Computer Aided Surgery*, **12**, 335-346.

Love, B., Jones, M., Tomlinson, M., & Howe, M. (2008). Predicting information needs: Adaptive display in dynamic environments. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 875-880). Austin, TX: Cognitive Science Society.

MacLin, O., & MacLin, M. (2005). Coding observational data: A software solution. *Behavior Research Methods*, **37**, 224-231.

Malik, R., White, P., & Macewen, C. (2003). Using human reliability analysis to detect surgical error in endoscopic DCR surgery. *Clinical Otolaryngology & Allied Sciences*, **28**, 456-460.

Megali, G., Signigaglia, S., Tonet, O., & Dario, P. (2006). Modelling and evaluation of surgical performance using hidden Markov models. *IEEE Transactions on Biomedical Engineering*, **53**, 1911-1919. doi:10.1109/TBME.2006.881784

Mehta, N., Haluck, R., Frecker, M., & Snyder, A. (2002). Sequence and task analysis of instrument use in common laparoscopic procedures. *Surgical Endoscopy*, **16**, 280-285. doi:10.1007/s004640080009

Neumuth, T., Czygan, M., Strauss, G., Meixensberger, J., & Burgert, O. (2009). Computer assisted acquisition of surgical process models with a sensor-driven ontology. In *MICCAI Workshop on Modeling and Monitoring of Computer Assisted Interventions (M2CAI)*, London.

Neumuth, T., Durstewitz, N., Fischer, M., Strauß, G., Dietz, A., Meixensberger, J., et al. (2006). Structured recording of intraoperative surgical workflows. In S. Horii & O. Ratib (Eds.), *SPIE Medical Imaging 2006—PACS and imaging informatics: Progress in biomedical optics and imaging* (Vol. 7, p. CID 61450A). Bellingham, WA: SPIE Press.

Neumuth, T., Jannin, P., Strauss, G., Meixensberger, J., & Burgert, O. (2009). Validation of knowledge acquisition for surgical process models. *Journal of the American Medical Informatics Association (JAMIA)*, **16**, 72-80. doi:10.1197/jamia.M2748

Neumuth, T., Strauss, G., Meixensberger, J., Lemke, H.-U., & Burgert, O. (2006). Acquisition of process descriptions from surgical interventions. In S. Bressan, J. Küng, & R. Wagner (Eds.), *Proceedings of the 17th International Conference on Database and Expert Systems Applications (DEXA '06)* (Vol. 4080, pp. 602-611). Berlin: Springer.

Neumuth, T., Trantakis, C., Riffaud, L., Strauss, G., Meixensberger, J., & Burgert, O. (2009). Assessment of technical needs for surgical equipment by surgical process models. *Minimally Invasive Therapy & Allied Technologies*, **18**, 841-849.

Padoy, N., Blum, T., Essa, I., Feussner, H., Berger, M., & Navab, N. (2007). A boosted segmentation method for surgical workflow analysis. In R. Larsen, M. Nielsen, & J. Sporring (Eds.), *Proceedings of Medical Image Computing and Computer-Assisted Intervention—MICCAI 2007: Part I. Lecture notes in computer science* (Vol. 4791, pp. 102-109). Berlin: Springer.

PostgreSQL Global Development Group (2009). *PostgreSQL: The world's most advanced open source database*. Retrieved July 27, 2009, from www.postgresql.org.

Racer Systems (2009). *RacerPro—Renamed Abox and Concept Expression Reasoner*. Retrieved September 28, 2009, from www.sts.tu-harburg.de/~r.f.moeller/racer.

Raimbault, M., Morandi, X., & Jannin, P. (2005). Towards models of surgical procedures: Analyzing a database of neurosurgical cases. In O. Ratib & S. Horii (Eds.), *Medical Imaging 2005: PACS and imaging informatics* (Vol. 5748, pp. 97-104). Bellingham, WA: SPIE Press.

Richards, C., Rosen, J., Hannaford, B., Pellegrini, C., & Sinanan, M. (2000). Skills evaluation in minimally invasive surgery using force/torque signatures. *Surgical Endoscopy*, **14**, 791-798.

Rosen, J., Brown, J., Chang, L., & Hannaford, B. (2006). Generalized approach for modeling minimally invasive surgery as a stochastic process using a discrete Markov model. *IEEE Transactions on Biomedical Engineering*, **53**, 399-413.

Sarkar, A., Dutta, A., Dhingra, U., Dhingra, P., Verma, P., Juyal, R., et al. (2006). Development and use of behavior and social interaction software installed on Palm handheld for observation of a child's social interactions with the environment. *Behavior Research Methods*, **38**, 407-415.

Schuster, M., Wicha, L., Fiege, M., & Goetz, A. (2007). Utilization rates and turnover times as indicators of OR workflow efficiency. *Der Anaesthesist*, **56**, 1060-1066.

Stanford University (2009). *The Protégé Ontology Editor and Knowledge Acquisition System Version 3.4.1*. Retrieved July 24, 2009, from http://protege.stanford.edu/.

Strauss, G., Fischer, M., Meixensberger, J., Falk, V., Trantakis, C., Winkler, D., et al. (2006). Workflow analysis to assess the efficiency of intraoperative technology using the example of functional endoscopic sinus surgery. *HNO*, **54**, 528-535. doi:10.1007/s00106-005-1345-8

Workflow Management Coalition (1999). *Terminology and glossary* (Document WFMC-TC-1011, Document Status–Issue 3.0). Winchester, U.K.: Author.