Estimation of Operator Input and Output Workload in Complex Human-Machine-Systems for Usability Issues with iFlow

Stefan Pfeffer[1], Patrick Decker[1], Thomas Maier[1], and Eric Stricker[2]

[1] Institute for Engineering Design and Industrial Design, Research and Teaching Department
Industrial Design Engineering, University of Stuttgart, Germany
{stefan.pfeffer,thomas.maier}@iktd.uni-stuttgart.de
[2] Center for Patient Safety and Simulation, University Hospital Tuebingen, Germany
eric.stricker@tupass.de

Abstract. Usability studies often use methods focused on product parameters. Test designs are processed in laboratories and evaluation is commonly performed by expert opinions. For validation studies we want to point out the importance of field studies and user and system oriented evaluation. For this purpose we want to present the methodological approach iFlow (information flow) as multiple assessment technique for usability issues in real or quasi-real (simulated) situations. The idea of iFlow is to assess input and output workload via video and audio recordings combined with subjective and objective measurement techniques of workload. In this contribution the iFlow method and an evaluative study in anesthesiology are presented. The added value to already existing methods and approaches is considered in the sensitivity of iFlow to identify situations of overload in a descriptive way. For design interventions it would be helpful to consult the iFlow chart to deduct cause and effect relations.

Keywords: Information Flow, Usability, Input Workload, Output Workload.

1 Introduction

Today the concept of usability is applied in many areas of human-machine-systems (HMS) expanding from its software ergonomic roots [1] to a broad understanding of any kind of interface between biology (human) and technology (cultivated products). In socio-technical-systems usability also considers networking of humans among each other which is mostly deficient in terms of efficient and error-free messaging. Because of the large amount and variety of information flow an operator has to handle in such systems, they are called complex.

An approach to operationalize this complexity in a macroergonomic perspective has been made by Manser [2] via the density of simultaneous and sequential processes in a time line analysis of anesthesia administration. Our methodological approach iFlow transfers this idea into the area of microergonomics, towards the estimation of operator input and output workload as criteria for usability considerations in complex

D. Harris (Ed.): EPCE/HCII 2013, Part I, LNAI 8019, pp. 167–176, 2013.

HMS [3]. Figure 1 shows the abstraction of such an HMS, the parameters that determine operator workload in the iFlow approach and the methods we used for assessment at a glance.

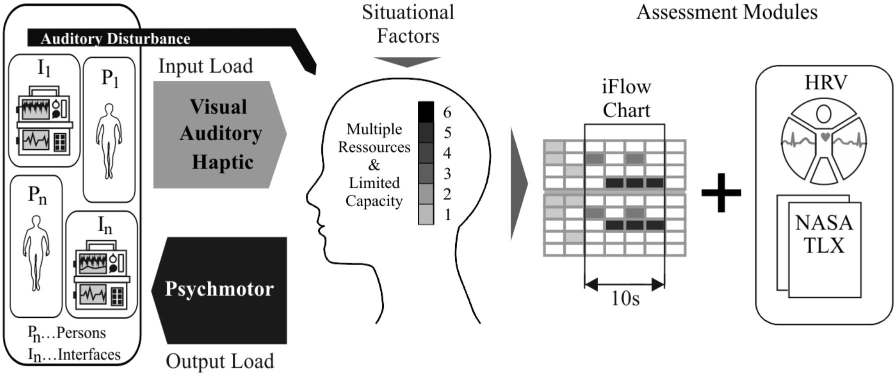

Fig. 1. iFlow model of a complex HMS and connected assessment parameters of iFlow

After a brief summary and discussion of typically used methods for usability assessment in chapter 2 the theoretical background (chapter 3.1) and the procedure of descriptive workload evaluation (chapter 3.2) are described. To show the added value of iFlow to already existing methods and approaches we want to present the results of an evaluative usability study in anesthesiology at the Center of Patient Safety and Simulation (University Hospital Tuebingen, Germany) in chapter 4.

2 Usability Assessment in Complex Human-Machine-Systems

The basic idea of usability engineering has its roots in human factors science. For this reason methods for evaluation arise from this discipline. Usability methods commonly are classified into formative (inductive) and summative (deductive) evaluation [4]. This division regards the method´s operational date – during or after design process. In human factors sciences methods are categorized into analytical and empirical methods [5], regarding the type of procedure. Besides these categorizations there are several other approaches to classify the large variety of methods (e.g. subjective/objective, descriptive/predictive, laboratory/field, qualitative/quantitative, static/dynamic). Each type of method can be beneficial in its own specific way. Human factors researchers have agreed that a multiple method approach would be the best way for usability assessment [6]. In practice it can be found that usability observations show a main focus in product oriented, static laboratory methods with qualitative outcome from subjective expert or user sources. This may be because of benefits regarding resource consumption (users, time, equipment).

Validation can be regarded as special kind of usability study. For example, in medical device design usability has to be validated according to the standards. Therefore the usability of the medical product has to be evaluated with high external validity and documented in a usability engineering file with comparable quantitative data. Hence we recommend that usability validation studies should be user and system oriented, performed under real (field) or quasi-real (simulation) dynamic conditions with quantitative outcome from an ideally objective and theory based assessment. With respect to a multiple method approach, subjective (user questionnaire) and psychophysiological assessment techniques should complete a validation study.

Especially the assessment of dynamic processes under real or quasi-real conditions is necessary for external validity. Moreover usability for validation shouldn´t be fixed only to observation of errors or nearly errors in user trials (product oriented, expert based) and subjective user opinions (product oriented, user based). Rather the assessment approach for validation should be extended and human factors methods for user oriented evaluation should be used more often. For example assessment of mental workload and situation awareness under real or quasi-real conditions could give a view on system usability and human reliability with respect to stressful situations. In the following chapters we want to present the iFlow method as an approach for estimating operator input and output workload and an evaluative usability study in a dynamic, quasi-real (simulated) anesthesiology setting.

3 The iFlow Method

The iFlow method is a descriptive usability assessment approach which is based on data collection of information input and output flows of the operator. This data is extracted of video and audio recordings and has to be transferred into the iFlow chart which is the basis for further evaluation steps along the timeline. The objective of getting quantitative data out of descriptive analyses is achieved by observing the density of input and output information flows (iFlows) by means of action codes similarly to Ekman´s FACS method [7]. This value in turn determines the level of workload. Weighted density analysis is combined with an objective (heart rate) and subjective (NASA TLX [8]) assessment method (see Figure 1).

3.1 Theoretical Background

The iFlow method has been developed on the basis of Wickens´ Multiple Resource Theory (MRT) [9]. It assumes that the external factors of the HMS, i.e., the interfaces have a main influence on operator workload. The theory also considers these external factors in the context of situational disturbance and control variables. The iFlow model (Figure 1) does not include internal factors like the current emotional state of the operator. In MRT input and output modules have disjunctive resource capacities. Hence each input (visual, auditory, haptic) and output (psychomotor) channel should not be overloaded singularly. In addition we want to recommend a limit for input (visual+auditory+haptic) and output load (motor+verbal). The division into input and output workload with regard to the user as sender and recipient of information flow can help to detect bottlenecks in both areas and specify deficits in interface elements of output (e.g. displays) or to interface elements of input (e.g. controls). Like limitations empirically found in anthropometry concerning physical workload (e.g. for raising a weight four times an hour) it would be a landmark for usability engineering to assess the amount of information a human operator can receive, process and turn into safe and efficient actions.

3.2 Workload Assessment Process

The basic idea of iFlow is the evaluation of time-related information density and the descriptive deduction of workload levels. Information flows in all channels of input and out are evaluated with respect to their potential degree of load for the operator. Classes of load have been developed after the VACP model [10].

VACP originally is used in a discreet, function-oriented way to predict workload for visual (V), auditory (A), cognitive (C) and psychomotor (P) channels in cumulated 10s intervals as well as an overall workload rate. Each channel or workload component is described by a 7-point ordinal rating scale (except auditory) [10]. The scales have been extended (second visual scale and kinesthetic scale) and further developed to an interval scale by pair comparison survey among pilots [11]. Also task time required for each task was considered for prediction of workload.

In order to evaluate the density of iFlow we used this basic work and designed new classes for a descriptive and continuous use. Several adaptions had to be made with regard to reliability and objectivity of behavior observation. After a recapitulation of the 7-point scales, ordinal 4-point (input) and 5-point (output) scales were formed with descriptors that are well distinguishable in video and audio recordings (see Table 1). For this reason, action codes have been developed to ensure a correct and consistent allocation of information flows to the classes. The interval scales of Bierbaum et al. (1989) refer to a sample of 20 pilots wherefore we decided to use an ordinal scale (following the order of the interval scale) showing more general applicability. Furthermore, we added the channel of haptics (H) and left cognition descriptors (C) out of account. The psychomotor scale was newly introduced so that a value of 5 (overload) is defined by a multiple task that is performed (two or more actions with different goals at once).

The limit of each workload component is fixed by Bierbaum et al. (1989) to the value that exceeds the maximum on any of the scales. Thus, a value of 8 implies a component overload and those channels can be identified that are associated with overload for each task and in total. We applied this systematization to generate specific limits of workload for input (V+A+H → Limit 8) and output (P → Limit 4) channels with the objective to detect situations of high workload that have an influence on the usability and safe use of the devices. Overload for input and output modalities can be described by adding up all information flows that come up in a continuous T=2-sec interval. Considering all input and output parameters in a socio-technical-system should lead to a higher external validity.

Table 1. Workload scales and limits of input and output referring to Bierbaum (1989)

Input (Limit 8)	**VISUAL**			
	No.	**Description**	**Action Code**	**Weight**
	V1	Visual Detection	Gaze <=2s	1
	V2	Visual Discrimination	Gaze >2s, Static, Target-oriented	2
	V3	Visual Tracking	Gaze >2s, Dynamic, Target-oriented	3
	V4	Visual Read, Searching, Orienting	Dynamic, Visually high attentive	4
	AUDITORY			
	No.	**Description**	**Action Code**	**Weight**
	A1	Auditory Detection	Digital Signal, Sound	1
	A2	Auditory Verification	Auditory Feedback	2
	A3	Auditory Decoding	Speech, Semantic Content	3
	A4	Auditory Interpretation	Sound patterns, Auditory high attentive	4
	Auditory Disturbance		Background noise (non-directive)	each 1
	HAPTIC			
	No.	**Description**	**Action Code**	**Weight**
	H1	Haptical Activation	Touch, Hold	1
	H2	Haptical Detection	Passive (Pressure), Active (Use Object)	2
	H3	Haptical Scanning	Active: Feel out, Haptic Feedback	3
	H4	Haptical Interpretation	Passive: Patterns	4
Output (Limit 4)	**PSYCHOMOTOR**			
	No.	**Description**	**Action Code**	**Weight**
	P1	Discrete Actuation, Speech, Walk	e.g. Push Button, Talk, Walk	1
	P2	Continuous Adjusting	Unimanual	2
	P3	Symbolic Production	e.g. Writing	3
	P4	Convergent Multiple Operations	>=2 Extremities, 1 goal	4
	P5	Divergent Multiple Operations	>=2 Extremities, >=1 goal	5

4 Evaluative Study in Anesthesiology

To evaluate the possibilities and limitations of iFlow for usability evaluation we examined a critical incident situation in anesthesiology. The anesthesia work system can be classified as complex human-machine-system. For a high external validity in validation studies it would be essential to test the products in context of the socio-technical-system with regard to situational factors (e.g., critical incidents). Because of safety reasons such situations can be produced by simulations. At the Center for Patient Safety and Simulation (Tuepass) it is possible to simulate such critical incidents. Beside the main purpose of crisis resource management the simulation can excellently be used for quasi-real usability studies [12]. Tuepass has four simulation rooms which can be established as operating or intensive care rooms. All devices are functional and the patient is represented via the full-scale simulation doll which has manifold options like pathological and physiological cardiac and respiratory sounds, difficult airway, centrally and peripherally palpable pulses, pupil reaction, chest drainage system etc.

4.1 Test Procedure

We chose an intensive care setting for the evaluation. Eight subjects (consulting and ward physicians) performed a critical incident situation in intensive care. Three cameras video- and audiotaped the observation. The subjects had to wear a bipolar one-channel ECG. After the observation they completed the NASA TLX questionnaire (weighting and rating). The observation lasted 5 minutes for each subject. The scenario began in the intensive care room with a critical incident noticed by the nursing staff and residents. The observation started when the subjects came into the running scenario.

4.2 Data Analysis

The audio and video data were analyzed by means of iFlow and transferred into the iFlow chart. Figure 2 shows a 2:40 min extract of the iFlow chart (subject 1). The iFlow chart consists of 4 sections (input, output, heart rate (HR) and load evaluation. The input and output sections are divided into the workload scale descriptors of Table 1. Descriptor P1 was itemized into P1-T (talk and discrete actuation) and P2-W (walk) for better overview. The information flows were manually extracted of the footage by the action codes. When a descriptor appeared twice in a 2-sec interval the values were added (e.g. 2x auditory decoding → 6). The time bars are coded from bright (value 1) to dark (value 6). The bar graphs in the HR section show the mean heart rate and standard deviation for a 10-sec interval. The graph illustrates the absolute deviation of heart rate in 1/min from the measured resting value of the subject. The load section is divided into input and output load showing the added values for each 2-sec. interval. Those values which exceeded the maximum were marked black. Hashes represent values >9. Sections of reduced video and/or audio data quality are shaded in grey. Accuracy of evaluation is lower in these sections.

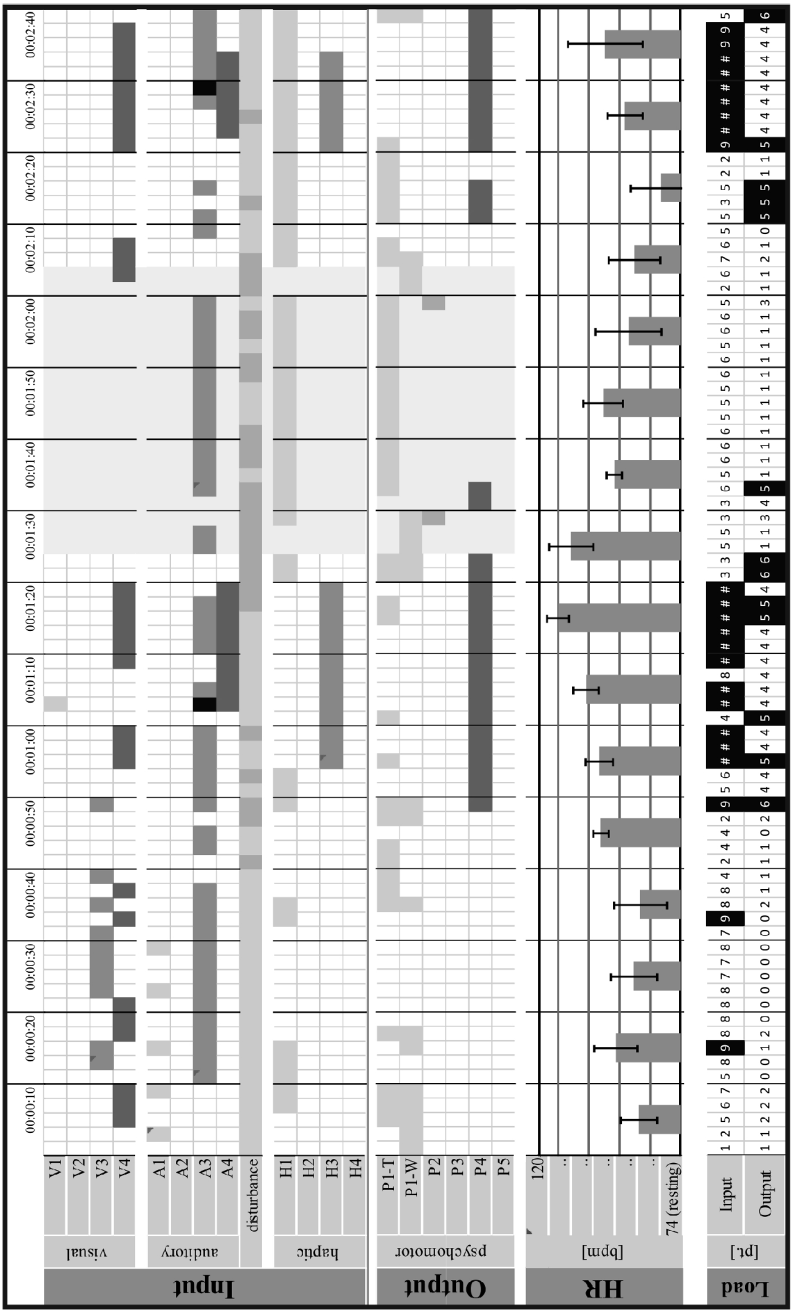

Fig. 2. iFlow chart excerpt (rotated through 90° anti-clockwise)

4.3 Results

Several devices were used in the scenarios by the subjects (patient monitoring system, defibrillator, bag valve mask, ultrasound system, syringe pumps etc.). The analysis did not focus on a specific device but rather on human-machine-interaction in general. The physicians were all familiar with the standard devices they were using in practice. That is probably why no errors in handling and operating occurred. Nevertheless, several near-errors could be observed (e.g., missed commando nearly leaded to shock the subject by defibrillator). In these situations an input overload (iFlow value >=9) could be observed while heart rate did not show any stress evidence.

Table 2 shows the results of the three assessment techniques. Mental and physical demand in original NASA TLX ratings are listed separately apart from the Overall Weighted Workload Score (OWWS). For each 5 min observation the average of iFlow input and output workload was calculated. Heart rate is listed for each subject as percentage deviation from the individual resting value for the observed timeframe.

Table 2. Results of the three assessment techniques

	NASA TLX [pt.]			**Workload [pt.]**		**Heart Rate [1/min]**
	Mental Demand	Physical Demand	OWWS	Input	Output	
Subject 1	90	30	65,33	6,83	2,39	+27,2 %
Subject 2	75	60	69,3	5,25	3,69	+49,9 %
Subject 3	65	20	45,67	5,23	1,85	+ 15,7 %
Subject 4	95	15	67,67	5,38	2,77	+38,2 %
Subject 5	85	20	68,3	3,7	1,83	+38,5 %
Subject 6	65	5	35,3	3,24	0,79	+27,6 %
Subject 7	75	10	48,3	4,45	0,77	+15,0 %
Subject 8	90	25	77,6	4,38	0,71	+36,3 %

We set up a correlation matrix (see Figure 3) for iFlow input workload and NASA TLX mental demand (hashes) as well as for iFlow output workload and NASA TLX physical demand (squares) although there are little differences in the constructs of workload. It is visible that the analyzed values by iFlow show dependence to the NASA TLX original ratings of the subjects in the two considered dimensions. OWWS and heart rate show inconsistent results towards the overall load of the subjects in this kind of data analysis.

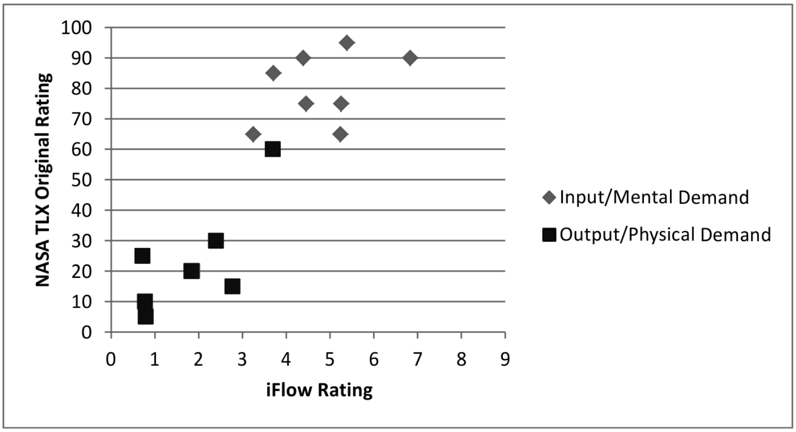

Fig. 3. Correlation matrix of workload parameters evaluated with iFlow and NASA TLX

5 Conclusion

The evaluative study shows both possibilities and limitations of iFlow. The claimed methods´ focus for validation studies (see chapter 2) of user-oriented and system-oriented usability studies performed in real or quasi-real situations under dynamic conditions can optimally be observed by video and audio recordings. The idea to quantify the qualitative video and audio data via action codes is realized by iFlow. In consideration of device design optimizations the strength of iFlow lies in a cause and effect analysis of single overloaded situations. The workload values for input and output channels produced by iFlow have to be considered as values of orientation. Even if not every situation in the evaluative study showed a near-error, these situations of input and/or output overload may potentially lead to an error. Passing just these situations without loss of efficiency or errors would mean a successful validation of the device. Approaches of design should gather interaction phases of overload and underload (monotony) to improve information-design and interaction-design. We provided a theory of action codes keeping objectivity of evaluation and rater reliability in mind (see Table 1). This theory has to be adapted in further steps (e.g. the distinction of auditory interpreted semantic content with and without situation awareness). Another benefit of iFlow is the possibility to document usability evaluations via the iFlow chart and discuss the results together with subjects.

It has to be stated that iFlow can´t and should not be used as single method. Following a multiple method approach iFlow has to be combined with subjective (questionnaires) and objective (psychophysiological) methods. The potential of psychophysiological methods lies in a continuous and time dependent data collection. In further studies we will consider continuous evaluating questionnaires too to get an attribution to specific phases. The chosen 10-sec. intervals show an adequate averaging of heart rate data (see standard deviations in Figure 2) in contrast to the overall heart rate means presented in Table 2.

In further studies it would be interesting to add EEG measurement for interpreting the cognitive readiness of the subjects by analyzing the alpha waves. This would probably give hints for the so far unconsidered module of cognition (C) and could help to deduct subjects´ mental states.

Concluding the added value of iFlow can be described as the possibility to quantify real-time data of human-machine-interactions within context of other process variables embedded in an overall situation considering user and system oriented dimensions of usability. In our evaluative study the calculated input and output workload is correlated with the NASA TLX dimensions of mental and physical demand. This result made us confident that a descriptive evaluation of workload limits as preliminary phases of errors in field studies are within reach.

References

1. Foley, J.D., Van Dam, A.: Fundamentals of Interactive Computer Graphics (Systems Programming Series). Addison-Wesley, USA (1982)
2. Manser, T., Wehner, T.: Analysing action sequences: Variations in action density in the administration of anaesthesia. Cognition, Technology & Work 4(2), 71–81 (2002)
3. Pfeffer, S., Maier, T.: Systematical improvement of an anesthesia workstation considering physical and mental workload. In: Salvendy, G., Karwowski, W. (eds.) Advances in Human Factors and Ergonomics 2012: Proceedings of the 4th AHFE Conference, July 21-25. CRC Press, Boca Raton (2012)
4. Sarodnick, F., Brau, H.: Methoden der Usability Evaluation: Wissenschaftliche Grundlagen und praktische Anwendungen. Verlag Hans Huber, Bern (2011)
5. Stanton, N.A., Salmon, P.M., Walker, G.H., Baber, C., Jenkins, D.P.: Human Factors Methods: A Practical Guide for Engineering and Design. Ashgate, Farnham (2005)
6. Vidulich, M.A., Tsang, P.S.: Mental Workload and Situation Awareness. In: Salvendy, G. (ed.) Handbook of Human Factors and Ergonomics, pp. 243–273. John Wiley & Sons, Hoboken (2012)
7. Ekman, P.: What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS). Oxford University Press (2005)
8. Hart, S.G., Staveland, L.E.: Development of NASA-TLX: Results of Empirical and Theoretical Research. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload, pp. 239–250. Elsevier, Amsterdam (1988)
9. Wickens, C.D.: The structure of attentional resources. In: Nickerson, R. (ed.) Attention and Performance VIII, pp. 239–257. Lawrence Erlbaum, Hillsdale (1989)
10. McCracken, J.H., Aldrich, T.B.: Analyses of selected LHX mission functions: Implications for operator workload and system automation goals. Anacapa Sciences, Fort Rucker, Alabama (1984)
11. Bierbaum, C.R., Szabo, S.M., Aldrich, T.B.: Task Analysis ofthe UH-60 Mission and Decision Rules for Developing a UH-60 Workload Prediction Model, vol. 1: Summary Report. Anacapa Sciences, Fort Rucker, Alabama (1989)
12. Stricker, E., Pfeffer, S., Trick, M., Rall, M., Maier, T.: Standardized research onergonomics, usability and workflow analysis using high fidelity simulation labs. In: Dössel, O. (ed.), Biomedical Engineering, SI-1 (2012) ISSN 1862-278X