

Towards Task Analysis Tool Support

Suzanne Kieffer¹, Nikolaos Batalas², Panos Markopoulos²

Université catholique de Louvain¹
Louvain School of Management
Louvain-la-Neuve, Belgium
suzanne.kieffer@uclouvain.be

Eindhoven University of Technology²
Industrial Design
Eindhoven, The Netherlands
{n.batalas, p.markopoulos}@tue.nl

ABSTRACT

This paper discusses challenges in contextual task analysis and the need of tools that support analysts to collect such information in context. Specifically we argue that the analysis of collaborative and distributed tasks can be supported by ambulatory assessment tools. We illustrate how contextual task analysis can be supported by TEMPEST, a platform originally created for experience sampling and more generally, longitudinal ambulatory assessment studies. We present a case study that illustrates the extent to which this tool meets the needs of real-world task analysis, describing the gains in efficiency it can provide but also directions for the development of tool support for task analysis.

Author Keywords

Contextual task analysis, ambulatory assessment methods, tool support.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI); Miscellaneous.

INTRODUCTION

Contextual Task Analysis (CTA) is a challenging, time-consuming and repetitive Usability Engineering (UE) task. Firstly, CTA is challenging because analysts have to venture out in the users' work environment and quickly adapt to fit their work, their culture, their jargon, etc., in order to understand and learn about the users, their tasks, and their environment. The most laborious part of it probably lies in capturing data regarding task performance when these are performed for real, such as common user errors, breakdowns, and workarounds, or ways in which the task overtaxes human capabilities. This means that analysts have to strike a balance between encouraging users to thoroughly demonstrate how they actually perform their tasks, and preventing users to explain the formal guidelines and procedures prescribed in the organization.

Secondly, CTA is time-consuming. While studies will differ a lot from each other, contextual observations and interviews, data transcription, their interpretation, and their organization into work models typically take one day's work per interviewee for a typical study; the overall CTA procedure ranging in duration from one week with a

small and focused project, to months with larger projects (Mayhew 1999).

Finally, CTA is repetitive. The purpose of CTA is to capture instances of a population of users while performing specific tasks. With three to five interviewees solicited per key user, it is not uncommon to have to interview more than 15 individuals (Mayhew 1999).

In this paper, we explore the potential benefits that lie in using Ambulatory Assessment Methods (AmA) to assist CTA with regard to contextual observations/interviews. AmA (also known as Ecological Momentary Assessment) is an umbrella term that encompasses a broad range of field methods that aim to capture data about processes in naturalistic settings, as they normally unfold in time. Such methods are applied in numerous fields and are known in several variants such as Experience Sampling, Diary Methods, Ambulatory Monitoring, etc.

In this paper, a signal-contingent sampling method is applied to gather typical information about the users and their tasks such as what, when, how frequently, what for, to whom, etc. The goal is twofold: to reduce the analyst's workload and to improve the contextual observations / interviews process.

The contribution of this work lies in that it identifies challenges for task analysis of distributed collaborative tasks, and sketches an approach by which analysis can be supported by AmA. Second, the paper provides a first, realistic account of the application of AmA in a field where, to the extent of our knowledge, no such work has previously been conducted. The reflection on a case study allows us to propose the method as beneficial to Contextual Task Analysis, and worthy of further elaboration.

The paper is structured as follows: the next section presents how AmA can help improve CTA by partially automating contextual observations/interviews. The tool support is then presented. This is followed by the presentation of a case study conducted in the sector of hot-dip galvanization. Questions raised during the study such as "why is task analysis difficult in this particular case?" or "how does AmA efficiently support task analysis?" are then discussed. The paper concludes by presenting some characteristics of AmA suitable as Task Analysis Tool Support.

RELATED WORK

Task analysis and task modelling have been researched extensively since the early days of Human Computer Interaction (HCI) research as means to represent users' tasks, related knowledge, and practices and use those as

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

OZCHI '14, Dec 2-5, 2014, Sydney, Australia

Copyright © 2014 ACM 978-1-4503-0653-9... \$15.00

the foundation of interactive systems design. A proliferation of models of users' tasks and related analysis methodologies were proposed at that time. Starting from hierarchical models of task performance (Annett 2004) derived through observations and interviews, the emphasis shifted first towards more cognitive models that represented user's knowledge of tasks (Johnson et al 1991), eventually moving to more ethnographically informed and contextualized models of users' tasks as in Contextual Design (Wixon et al 1990). Over the years substantial research effort has been invested into developing representations for task models that would be more or less amenable to editing and transformations and there has been a proliferation of research into software tools supporting model based/driven design of interactive systems, e.g., (Markopoulos et al 1992, Bodart et al 1995, Paterno et al 1997, Vanderdonck 2008). Task analysis and task modelling are by now mainstream and have even been included in standardized procedures for supporting interaction design, e.g., ISO standards (Bevan 2001), or widely read textbooks for usability engineering (Mayhew 1999).

Despite the maturity characterizing this field and the wide adoption of task analysis methods in industrial practice, it is perhaps surprising how little effort has been invested to develop tools supporting data gathering during field studies that inform task analysis. This process still relies largely on face-to-face interviews, participant observation, etc., and generic technologies for making video and audio records of observations and interviews. The present research was inspired by the needs of supporting task analysis in the field where contextual and resource constraints make it difficult to spend sufficient time in the work-floor or to spread analysis effort among several investigators at once. We examine the applicability of tools initially intended for Ambulatory Assessment to support task analysis procedures and methodological questions pertaining to their use.

Task Analysis: a process perspective

The literature on the process of task analysis has evolved from an ergonomics perspective, describing simple interviews and observations techniques, and was later extended with elements of knowledge elicitation and ethnography. To a large extent methodology focuses on what questions should be put to informants and in which way. E.g., Knowledge Analysis of Tasks (Johnson et al 1991) suggests the use of repertory grids to derive constructs that make up a user's understanding of a task domain, and laddering interviews regarding task decomposition and sequencing. Contextual Inquiry (Wixon et al 1991) and Contextual Task Analysis (Mayhew 1999) emphasize the social relationship between the analyst and the informant, the ethnographic procedures, the type of questions that can be asked, and the group procedures by which information obtained during field work can be synthesized into a representation shared with a design team. Methods like card sorting are sometimes mentioned in passing for identifying the hierarchical organization of task related knowledge (Mayhew 1999).

Despite their wide diversity and the different scientific paradigms underlying these methods that range from neopositivistic and Tayloristic views of work to ethnographic and socially constructed views of work organization, they share some common elements:

- While direct observation plays an important part in identifying breakdowns, workarounds and resources people use to carry out their tasks, the main source of information regarding sequencing, information flows, workload, etc. remains the informant who is questioned directly about them.
- Analysts and informants are co-present which costs time and effort to both parties. Especially for collaborative and distributed tasks observation and interviewing require the involvement of multiple analysts or the use of video recording and retrospective analysis. Even then, the information needed to populate task models is very hard to identify (Pinelle et al 2003).
- Collecting task related information is not supported by tools. Card sorting is perhaps one of the few exceptions where practical and industrially proven tools have been developed to facilitate the work of the interaction designer (Hudson 2013) but such sessions are held outside of the work context, and decoupled from observations.

In conclusion, there is a lack of tools to support task analysis activities in the field which remains labour intensive, requiring face-to-face interaction, and often relies upon old fashioned pen and paper, or at best audio/video recording and transcribing to capture information. These media are very limited and restrictive for a number of situations:

- Where multiple distributed workers collaborate.
- Where direct observation is difficult or impossible.

We argue that in practical work, task analysis concerns such situations as often as not. Where fast-paced tasks are involved, video observation and subsequent video analysis are applicable, e.g. using the Interaction Analysis Laboratory (Jordan et al 1995) method. Still these do not allow real-time inquiry and observations are not sufficient or appropriate for understanding the thoughts and experiences of people observed, and they take a long time to execute: four times the duration of the video logs (Jordan et al 1995).

Ambulatory Assessment for task analysis

Ambulatory Assessment (AmA) makes use of field methods to assess the ongoing behaviour, knowledge, experience and environmental aspects of people during task execution in their natural setting. It is popular in clinical psychology for capturing ecologically valid data from subjects as daily life unfolds, and for reducing retrospective biases (Trull et al 2013). The tools that

support AmA methods include but are not limited to repeated-entry diaries, experience sampling and ecological momentary assessment, monitoring of physiological function, and acquisition of ambient signals. AmA methods can provide a complementary approach to conduct task analysis. In the field of HCI these methods were introduced by (Rieman 1991) who advocated diary methods for soliciting information close to the moment and in the context to which questions might refer. Diaries are used extensively and can be supported by multi-modal data capture (Carter et al 2005).

More recently interest in Experience Sampling (Larson et al 1983) has grown as the focus of HCI has shifted away from a usability focus and a task centric view on people's activities, to more holistic understandings of user experiences in the context of daily life and work. Related tools and methodological variations have also attracted considerable interest, e.g., (Khan et al 2008), but computerized user sampling technology has not yet been used for task analysis.

Especially for the cases discussed above involving distributed collaborative tasks, there is opportunity to support task analysis with data collection in situ, by way of Ambulatory Assessment methods. With AmA tools, information is captured in real time, and participants can provide accounts that relate to the present moment and their immediate experience, within the actual environment where the experience takes place. Moreover, recent developments in smartphone and web technology meant that the collection of data can be performed economically, both with respect to time and resources and in a fashion that is either non-disruptive or minimally disruptive to the activity that is taking place. Thus, it is possible to get ecologically valid data and minimize retrospective biases.

Within task analysis in particular, such a venture requires narrowing down information collected to the interests of the task analyst and automating the logistics pertaining to triangulating inputs from multiple workers carrying out collaborative and shared activities. Further, it is necessary that the analyst can easily set up computerized experience sampling and diary procedures, a task that has been traditionally a very demanding software engineering effort in the context of scientific research.

METHOD

The method proposed to conduct tool-supported contextual task analysis is divided into three steps: (1) Task model hypothesis, (2) Ambulatory Assessment (AmA), and (3) Contextual observations/interviews.

Task model hypothesis

This step focuses on the analysis of the formal procedures and artefacts in order to produce a task model hypothesis which will drive the setting of questions and experimental design for the AmA step.

Questions need to be contextualized to be relevant to people's tasks and current context (Markopoulos et al 1992) In order to contextualize questions, a *task model hypothesis* (in analogy to Cooper's persona hypothesis

(Cooper et al 2007)) has to be constructed. This can be an early cut of the task model that is obtained by interviews, or by examining procedure manuals and artefacts made available by the organization concerned. It is well known that such normative descriptions of user tasks are a far cry from actual practice. They nevertheless do provide a way of narrowing down the focus of inquiry at a particular moment in time and a particular user and context. A task model hypothesis will relate contexts (time, place, current activity) to different questions that the analyst needs to answer.

Examples of how the task model hypothesis can guide the sampling protocol are the following:

- When a team leader initiates a task, queries for what tasks users are engaging in can be initialized with possible tasks for each person.
- Tasks on the work floor are typically tied to a specific location and responsibility. Devices used for in situ data collection can be personal or location bound enabling a selective presentation of queries.

Ambulatory Assessment

In situ data collection takes places during this step. Data about the users (e.g. level of expertise in performing specific tasks), statistics about the task (e.g. relative frequency), and opportunities for improvements (e.g., errors, problems, workarounds, breakdowns) are gathered. They provide a structure to conduct contextual observations/interviews by highlighting users' issues and concerns in the current work process to focus on.

By processing and adapting to answers inquiry can become more specific and context relevant:

- Answers to questions give rise to more questions: by monitoring responses, the analyst can adapt the questioning protocol. While not a real time interview, this still allows for probing questions, to obtain more information on context and experiences relevant to a reported event.
- When a user reports to be engaged in a specific activity or when sensing and computation helps make that inference automatically, the system can trigger questions that the analyst would normally ask face to face: e.g., why (laddering up the task hierarchy), how (laddering down to understand task decomposition), information flows and dependencies to co-workers, breakdowns and strategies of users. E.g., "describe how you make this task efficient", "describe how you ensure you do it right", etc.

Information on duration, location, of specific tasks or known bottlenecks can also be requested close to the moment of task performance.

The TEMPEST platform

TEMPEST (Batalas et al 2012) is a software platform aimed at orchestrating in situ data collection. It has been developed for the purpose of enabling analysts without necessarily having expertise in software engineering, to

easily develop and administer surveys in situ, typically using smartphones or other mobile devices. TEMPEST consists of several modules:

- A web-application that can render interfaces on multiple target platforms while preserving the essential interface content and structure.
- Native applications that envelop the web-application and that enables the specification of behaviour of devices carried by respondents.
- A GUI that allows configuration of the interfaces and the behaviour of the web-based and native clients, and is meant to democratize the effort of building protocols and monitoring data.
- A server for these components to retrieve configurations from and submit data to.

Data collection can take place in both online and offline situations. As data collection progresses, it is possible to monitor the process and to perform real time modification of the parameters of the data collection procedure.

Access to TEMPEST allowed us to capture, alongside responses from participants, also the time it takes respondents to answer the questions. Such data might indicate the feasibility of applying the AmA method realistically in a work setting. TEMPEST is one amongst a class of systems that share similar goals, and have been built to service AmA studies. We anticipate that other systems in the same class can perform suitably well.

TEMPEST was selected for the present study as it offers more flexibility and control allowing to easily set up customized sampling protocols and surveys, nest questionnaire items allowing for questions to be asked only at the appropriate context (e.g., depending on earlier answers or depending on context sensing). Offering a comparison on the suitability of specific applications to the purposes of task analysis is beyond the scope of this paper. Rather our aim is to expose the potential of the more general class of systems supporting AmA as aids for task analysis.

CASE STUDY

The case study concerns a task analysis carried out to inform the design of a bespoke software system that will be developed to automate an information management at the workplace. It took place in a company whose core-business is hot-dip galvanizing to produce rust proof steel sheet for the automobile industry. Hot-dip galvanizing consists of applying a zinc coating on steel products by immersing them in a bath of molten zinc in order to protect the underlying steel from corrosion. Nowadays, hot-dip galvanizing is operated on a continuous galvanizing line: the sheet is continuously fed through a cleaner, an annealing furnace, and then into a molten zinc bath. At the exit either the sheet is left to freeze as galvanize (GI) or immediately reheated by open-air burners as galvanneal (GA) (Dallin et al 2005).

Key tasks and key users

In order to ensure the quality of the production, workers are requested to follow procedures. Paper-based

checklists are the artefacts used to support the application of these procedures. Figure 1 presents the artefact used to prepare the production of exposed parts of vehicle, namely the “GIZ production”. The artefact is divided into four areas (from top to bottom): the headline, the workers involved, the checklist of items, and the space for signature. This checklist is a countdown of 30 items, each to be checked at a specific time by specific workers, before the GIZ production can start.

Title: GIZ - preparation Code: XX-AA-XXXXX			
Functions involved in its execution			
check	item	time	who
	item 1	09:00	x, y, z
	item 2	09:00	y, z
	item 3	09:15	y, z
	item 4	10:15	x
	item 5	10:45	x
	item 6	10:45	x, y, z
	item 7	11:00	x, y, z
	...	...	...
	item 30	12:30	z
Date and approval:			

Figure 1. Artefact used to prepare the GIZ production.

The completion of this checklist requires the collaboration of at least seven individuals: one team leader, one quality technician, four operators allocated alongside the line (entrance, bath, completion, and exit), and one mobile operator whose duties include both taking care of the furnace and supporting co-workers (Figure 2).

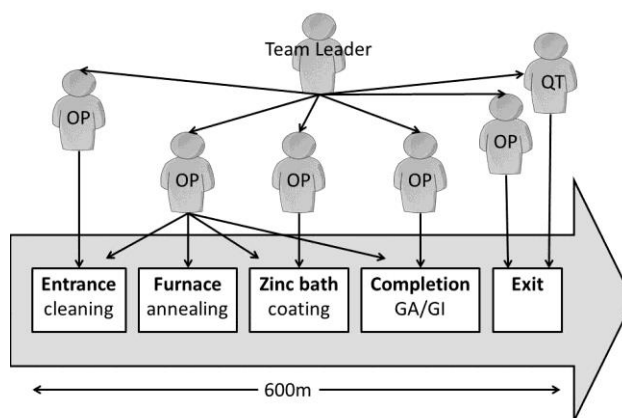


Figure 2. Distribution of co-workers alongside the line from entrance to exit. OP stands for operator and QT for quality technician.

Since the production is continuous, four teams work in rotating shifts of 8 hours: morning (A), afternoon (B), night (C), plus one team at rest. The workforce in production includes overall 41 persons: 32 operators, six quality technicians, and three team leaders. The

completion of this particular checklist was chosen as key task with managers and user representatives because the GIZ production is a critical and frequent activity at the company. In addition, its automation aligns with the identified business goals by supporting work processes effectively and efficiently.

Questionnaires and experimental design

Questionnaires and a study plan design were created during the task model hypothesis phase (Step 1). The purpose of the task analysis was to understand how workers actually think about, talk about, and execute the 30 items on the GIZ checklist. Based on the artefact a sequence of 12 questions to be asked systematically for each item was elaborated. Table 1 shows these questions.

To provide participants with context, a textual description of the item into focus was served as an announcement at the beginning of a 12-question sequence. This description was a copy of the one found in the original checklist, with a minor adaptation where one item was merged with its predecessor to which it referred.

Eventually, the representation of the checklist resulted in 29 sets of questions, each containing an announcement screen and a twelve-question set, apart from one of these (corresponding to the merged checklist items), which had an additional follow up question. The material produced within TEMPEST to carry out the investigation about the items included 42 screens overall: 29 announcement screens, 12 item-related screens, and one additional screen for the item 18 exception.

TEMPEST screens were presented one at a time to participants who were in possession of mobile devices (see Figure 3). The TEMPEST Android client notified the participants to read the announcement screen and answer the questions. The analyst configured the timing of these signals in the authoring environment of TEMPEST. Four mobile devices running the TEMPEST Android client were used: three Smartphones (Samsung Galaxy Ace, Samsung S3 Mini, Sony Xperia M) and one tablet (Samsung GT 7100).



Figure 3. An open question served to a participant.

In order to maintain a defensible external validity in the study, a stratified sampling of participants was made across posting (entrance, bath, furnace/mobile, exit, and team leader). Overall 12 participants were recruited, that is four per shift. These profiles were selected because they are the most involved in the execution of the checklist and because they have intimate knowledge of

the processes. The study lasted nine days, three to four items being tested every day. On the very first day, all 12 participants were given a small training session, and then the mobile devices were left in their possession. For the remainder of the experiment, at the end of his shift, each participant would hand the mobile device assigned to him over to the participant filling the same position in the next shift.

The total individual answers that could have been collected were 4188 (12 users x (12 questions x 29 items + 1 question item 18)). In the frame of such experiments, it was expected that some questions would not be answered, because of external factors. E.g., the signalling might take place at an inappropriate moment, or the participant's attention would be demanded elsewhere.

In situ data collection

In situ data collection took place at the factory during the Ambulatory Assessment phase (step 2). TEMPEST was the tool used for data collection and to generate ready-to-use data to analyse. Participants were provided with mobile devices with TEMPEST installed. Sequences of questions were triggered on these mobile devices (Figure 3) and answers were recorded in a database. Data such as time, shift, posting/function, question type, answer, and execution times were extracted from the log-files.

The Wi-Fi installation in the factory was not accessible from the production line. Instead, to keep monitoring the data collection in real time, mobile devices were equipped with 3G connectivity. This had the side effect of providing internet access to the workers carrying the phones, and in a few cases the quota of traffic allowed by the 3G provider was readily consumed for recreation.

Additionally, a few glitches in the software were not avoided, and there were instances when workers had been asked to restart the software running on the phones, to avoid data loss. Despite these glitches, 70% of the expected answers were collected, and overall the workers were prompt to incorporate the burden of participation into their work.

Data analysis

The analysis focused on the answers provided by participants in Table 1. First, our analysis indicates that the current user task organization model (i.e. the user model) is different from the one assumed by the analysis of the formal procedures (i.e. the task model hypothesis) in the following ways:

- Regarding familiarity with items (Q1): the task model hypothesis led us to believe that familiarity was defined strictly by the job functions assigned to items. The team leader appears to be the most familiar worker with any item, and his collaborators seem to be rather familiar with the items they are not in charge of.
- Regarding the timeline (Q3-4): the countdown between items indicated on the original paper checklist is almost never followed as is.

Question	Type
Q1. Please indicate your degree of familiarity with this task	Single-answer
Q2. How frequently is this task executed	Single-answer
Q3. Please indicate when it was executed for the last time	Date-answer
Q4. Please indicate when it will be executed next time	Date-answer
Q5. Please select in the list all the possible contexts where it takes place	Multiple-answer
Q6. Why does it have to be executed?	Open question

Question	Type
Q7. Please indicate a mean to facilitate or to improve this task	Open question
Q8. Please give an example of possible problem during its execution	Open question
Q9. Please give an example of error committed during its execution	Open question
Q10. Please select in the list all the participants to this task	Multiple-answer
Q11. Please indicate who asks for its execution	Multiple-answer
Q12. Please indicate to whom the related result is communicated	Multiple-answer

Table 1. The 12 questions asked for each item on the GIZ checklist. The options presented to the participants for answering single- and multiple-answer questions refer to the organizational culture; hence they have been omitted.

- Regarding the participants assigned to a task (Q10): the compliance of the work as it is effectively done with the formal procedures depends on items. The items related to the bath of zinc are the most physically challenging on the line and requires teamwork. One or two more workers than those assigned on the original paper checklist participate to these items.
- Regarding people who initiate and collect results (Q11-12): the task model hypothesis led us to believe that both initiators and collectors were the collaborators assigned to the execution of items. It turned out that it is the team leader who initiates over 85% of the items and collects over 90% of the results.

Second, our analysis indicates differences between teams in the way the collaboration is organized (Q10). For example, the implication of the team leader in the execution of tasks or the support provided by the mobile operator to his co-workers varies from one team to another.

Finally, our analysis indicates convergences:

- The reasons mentioned to execute the tasks (Q6) are distributed amongst three categories: following instructions (40%), cleanliness (30%), and quality of the production (25%).
- The means to facilitate/improve the tasks (Q7) are distributed amongst three categories: automation (60%), better care of the zinc bath (30%), and purchase of new equipment (10%).
- The problems that were mentioned (Q8) are distributed amongst two categories: technical problems resulting in a halting or slowing down of production, or those resulting in a downgrade of the coil (60%), and the accidents or inconvenience related to the work environment (40%).

- The errors that were mentioned (Q9) were mostly related to manipulation of the zinc bath (82%), the remaining 18% being attributed to lack of time and forgetfulness.

Contextual observations/interviews

The closing step of this method is when conventional contextual observations/interviews are conducted. The way they were conducted has been entirely driven by the findings of AmA. First and foremost, the focus of contextual observations/interviews was narrowed to what appeared as key locations and key functions: the zinc bath and the exit as key locations on the one hand, and the team leader, mobile and bath operators as key functions on the other hand. The other functions involved in the task have been only subjected to interviews.

The contextual observations/interviews were meant to confirm or dispel doubt regarding the findings extracted from AmA. The aspects into focus and elicited during interviews were:

- The communication flow between the co-workers (assignment, initiation, and collection of results)
- The countdown and the relative frequency of items (e.g. most of the items related to the zinc bath are executed no matter the type of production)
- The collaboration within each team
- How/why could automation help facilitating the tasks, even preventing problems or errors?

Following this data collection the analyst moved on to consolidate and organize results into task models. Task modelling is beyond the scope of this paper; hence the resulting task models are not presented here. However, task models and suggestions for improvement have been validated by the workforce through subsequent interviews and storytelling exercises.

DISCUSSION

Why is task analysis difficult in this particular case?

Traditional task analysis requiring observations and interviews is difficult in this particular case because of the work environment and the work organization. First, the work environment is unfriendly: no natural light, a lot of noise, dust and dirt, and a high risk of accidents. Consequently, both workers and visitors have to wear earplugs, safety shoes, helmet and coveralls at any time anywhere alongside the production line. This makes observations and interviews uncomfortable. Given that the execution of the GIZ checklist is spread over time, they become also exhausting, stressful, and inefficient.

Second, the work organization is complex: the execution of the task involves seven functions concurrently, the specific role incumbents change over time, and workers are distributed in space (Figure 2). Additionally, the production runs continuously 24 hours a day and follows rotating shifts. Due to this organization of work, traditional contextual observations/interviews would have put excessive demand on the analyst's time and resources, and yet spread in time. It would have taken up to six weeks full-time to the analyst to conduct contextual observations/interviews with half of the workforce involved in the task, on a one day's work per user unit (Mayhew 1999).

How does AmA support task analysis?

Questions were administered via TEMPEST on multiple locations, to different workers and with sensitivity to the task context. To perform the equivalent questioning process without TEMPEST, the analyst would have had to ask over 4000 questions overall, that is to ask over 300 questions to 12 different workers. This would have been embarrassing, tedious, or even impossible for both the analyst and the workers. The workload for a single human analyst would have been considerable and most likely resulted in unstructured and fragmented results.

In addition, the workers are under time pressure and primarily need to do their job. Any interruption from an analyst can be perceived as intrusive or as an unwanted distraction when controlling an industrial process or executing a physically demanding task. With TEMPEST, the workers were offered the possibility to snooze the triggers and therefore to postpone their answers to a most appropriate time for them.

How does AmA improve task analysis?

Table 2 presents the cost in hours of the contextual task analysis supported by TEMPEST. Overall, it cost 298 hours with a two-thirds one-third split between analyst and workers (see Table 2, 208 analyst hours versus 90 galvanizing company hours). The 38 user hours spent during in situ data collection were extracted from the log-files. To perform the equivalent questioning process with without TEMPEST would have at least doubled the costs of the questioning process: the current 38 hours spent by the users and at least 38 more hours used by the analyst to conduct the interviews. We believe that the dialogue between analyst and interviewee necessary during interviews (i.e., speaking the question out loud, processing the question, speaking the answer out loud,

processing the answer, logging the answer) should result in even more expensive costs. Clearly further research is needed to demonstrate the advantages of Task Analysis Tool Support (TATS) approach over traditional methods.

However, TATS reduced the analyst hours by 12% in this exploratory case study. The approach cost 12 analyst hours of TEMPEST use and enabled us to save 38 analyst hours of interview (see bold data in Table 2), that is 26 analyst hours saved. The benefits are even greater because of the possible gain from reuse. The reuse of the TEMPEST material produced during this case study is possible and would reduce the costs of task analysis for other populations/tasks (Bias et al 2005).

Usability Engineering Task	Analyst	Manager	IT	User
Interview team and users	10	6	2	2
Identify key actors and tasks	4	2		2
Analyse artefacts	16			
Draft questionnaire	16			
Program TEMPEST protocol	10			
Pilot questionnaire (2 analysts)	16	2	2	4
Revise questionnaire	2			
Revise TEMPEST protocol	2			
Recruit users	4	2		
Install and kick-off (2 analysts)	28		2	12
Run in situ data collection				38
Monitor data collection	4			
Analyse data	20			
Conduct contextual observations	16			8
Consolidate data	8			
Obtain task model	12	2		4
Documentation	40			
TOTAL	208	14	6	70

Table 2. Cost in hours of Task Analysis Tool Support.

Factor	Groups	Sample size	Mean	Variance
Shift	A	835	23.38	944.11
	B	888	18.22	586.76
	C	969	16.95	540.86
Posting	Leader	718	19.61	724.05
	Mobile	541	16.79	507.48
	Bath	757	18.48	720.96
	Exit	676	22.15	746.64

Table 3. Response time in seconds per shift and posting.

Factor	Sum of squares	df	Mean square	F	p
Shift	20248.64	2	10124.32	14.87	<.0001*
Posting	9440.01	3	3156.67	4.59	0.003*

Table 4. One-way ANOVA results on response time.

In addition, the automation of data collection eliminates mistakes such as forgetting questions, omitting items, losing track of users and answers, or making mistakes during transcription. It allowed detection of differences between workers/teams, which would have been difficult to notice in standard contextual interviews (e.g., the organization of collaboration within teams). Finally, during contextual observations/interviews, users are

asked to demonstrate their problems with the current system, workarounds, and the most common errors. Being aware of them beforehand thanks to AmA enables analysts to better encourage users to do so.

How does tool-supported task analysis also feed other Usability Engineering tasks?

We examined the time it took respondents to provide answers using TEMPEST. Outliers (i.e., either answers judged as irrelevant, or response time more than two standard deviations difference from the mean) were removed from the quantitative analysis. We analysed overall 2692 response times according to two factors: shift (A, B, C) and participant function (leader, mobile, bath, exit). Detailed results are presented Tables 3 and 4.

A comparison of performance between shifts indicates that fewer questions were answered during shift A than during shifts B and C (Table 3: 835 answers during shift A versus 888/969 answers during shift B/C). Similarly, participants were slower during shift A than during shifts B and C (Table 3: average response time of 23 seconds during shift A versus 18/17 seconds during shifts B/C). The night shift is the most performant. These differences are highly significant (Table 4: $F(2, 2692) = 14.87$, $p < .0001$). Similar results can be observed between postings (Table 4: $F(3, 2692) = 4.59$, $p = .003$).

Aside from the modelling phase, such statistics about the users, their tasks, and their environment obviously feed other Usability Engineering tasks:

- Prior to task analysis, user profile was conducted with the operators and revealed that the furnace operator is mobile and supports his co-workers. However, there was no evidence about his diminished availability and greater efficiency to perform tasks. Statistically demonstrated, these characteristics now directly feed into the user profile.
- Similar outcomes from AmA directly feed into the work environment analysis. E.g., shift A is the busiest of the day during which workers have to divide their attention between many activities. This results in fewer answers with slower response times than in shifts B and C (see Tables 3 and 4).
- Such statistics can guide rework engineering and design choices of other features (Montero et al 2005). They also can inform the setup of usability evaluation with operators. For instance, the statistics can drive the schedule and the duration of the evaluations, the location where they should take place, and the recruitment of participants.

User satisfaction

User satisfaction with the AmA procedures has been assessed through a satisfaction questionnaire where participants had to score their level of agreement with five statements on a 5-item scale (1=totally disagree, 2=disagree, 3=neutral, 4=agree, and 5=totally agree).

The five statements were formulated as follows:

- S1: The questions were clear.
- S2: The questions were easy.
- S3: Answering was fast.
- S4: The questions interfered with my schedule.
- S5: The number of questions was appropriate.

The responses indicated that participants were satisfied with the study. The questions were judged easy and fast to answer (S2 score=4.09; S3 score=3.81), rather clear and appropriate (S1 score=3.63; S5 score=3.63), and did not interfere much with their schedule (S4 score=2.81). In particular, the questions reasonably interfered with their schedule during the morning shift (S4 score during shift A=3.50 with participants' answers equally distributed between "neutral" and "agree"). On the contrary, they did not interfere with their schedule during the afternoon and night shifts (S4 score during shift B=2.67 and during shift C=2.25). These results are consistent with the quantitative results presented Tables 3 and 4.

Additionally –and surprisingly, although the questions were occasionally triggered at an inappropriate time and although the process of answering the very same questions for each item might have felt repetitive and boring, the participants were willing to contribute to the study and answered more questions than they thought they did. Most of the participants (10 out of 12) thought they answered between 15 and 30 questions a day, while they actually answered 36.

Real time monitoring

An open question dedicated to suggestions and comments closed the sequence of questions that was served as the last one of day 2 (for the record, day 1 was dedicated to the training). The purpose of this question was to get as soon as possible a spontaneous feedback about the participants' satisfaction with the study.

On the one hand, it has been particularly useful to modify slightly two questions the participants were struggling with (see Table 1 Q3 and Q4). Initially, these questions required to also indicate not just a date but also a time of execution. This time has been indicated as irrelevant by two third of the participants and as such was suppressed in these two questions.

On the other hand, suggestions to modify the time slots when sequences were served during the shifts A and C were made. Sequences were served to close to the end of these shifts. Sequences have been shifted accordingly one hour earlier.

TOWARDS TASK ANALYSIS TOOL SUPPORT

The case study presented demonstrates the feasibility of the approach advocated in this paper. Task analysis has been successfully supported by TEMPEST and the functionality it provides was instrumental for the case presented. There are a few key characteristics that make TEMPEST suitable as Task Analysis Tool Support (TATS):

- End user (analyst) configurability. Design projects cannot afford to invest software engineering effort

for surveying tools. It is important that non-software engineers are able to configure efficiently and independently complex sampling protocols, suitable for the context and population at hand. For such an approach to be applicable in different contexts, it is not possible to rely on a single device. It is important that the tool will work across platforms, to fit ergonomic and mobility requirements (Florins et al 2006), perhaps adapting to the informatics technology used in the task under study, or even relying on the devices that the respondents happen to own.

- Real time monitoring of responses and on the fly adaptation of the sampling protocol are critical. Task analysis performed in industrial contexts, is never done with the amount of forethought and planning that are possible and necessary in scientific research. It is therefore necessary to monitor the progress of the study and to adapt the protocol to the findings and situation at hand.

TEMPEST has not been created specifically for task analysis nor with the needs of the specific project or analyst in mind, which is promising for the generalizability of the approach presented. Nevertheless, it is valuable for future research to examine different cases as well. Specifically it is not clear whether a certain amount of structure and mutual dependencies should characterize the tasks studied before such tool support is found necessary. Further, it is important to find out how different analysts will respond to it, especially when they do not do it in the context of a research project. Note however, that in the case presented the adoption was driven by the need of the analysis rather than for testing the tool, but the expertise and willingness of the analyst to try new technology cannot be taken for granted outside a research context.

Context sensing (integrating sensor readings) and multi-modal information capture would largely extend the applicability of Ambulatory Assessment method as a TATS. While current technology and TEMPEST enables this, it is important to enable non-engineers to easily map context sensing to elements of a task and a sampling protocol. Extending TEMPEST or other in situ data collection tools to enable this seems an important step towards the development of TATS.

Over and above tool support, related methodology needs to be better understood. In this case our approach consisted in three main steps: formulating a task model hypothesis and developing a task and context dependent questionnaire and experimental setup, performing in situ data collection, and conducting contextual observations/interviews to complete the task analysis process. It is perhaps early to talk of shaping a general methodology for embedding TATS in the analysis process, but it is clear that one is needed. Formulating the task model hypothesis appears to be a key step. In the present case, this was expressed in artefacts and documentation as this was suitable for the task studied. To enhance the applicability of the method this hypothesis, is most generally going to be articulated using

a task modelling notation, e.g., TKS (Johnson et al 1991), CTT (Paterno et al 1997). It would be beneficial to support the link between a task model and the question asking protocol by procedural guidelines for the analyst and with software support. Such a link would facilitate the process of constructing the sampling protocol and mapping responses back to a task model. Further, it has to be shown in different application context how feasible it is for analysts to construct a hypothesis in sufficient detail to guide the information gathering procedure.

Finally, ambulatory monitoring is not void of methodological limitations. For example, the method may be experienced as intrusive in some contexts, it may bring about self-selection bias, or even priming effects issues that are common in several survey methods and can be accentuated by the specific characteristics of computerized experience sampling that was used in this study. These research considerations have traditionally concerned researchers planning field studies (Alliger et al 1993) and need to be taken into account in designing tools and developing methodology for task analysis.

CONCLUSION

This paper identifies a gap in the current research literature regarding the development of tools to support information gathering activities for the purposes of task analysis. It argued that for distributed collaborative tasks, use can be made of Ambulatory Assessment methods, provided that they support contextualized sampling of user attitudes and experiences, and real time monitoring of respondent feedback. A number of requirements to enable the practical use of these technologies in an industrial context have been mentioned, primarily to lower the threshold for setting up and administering in situ surveys. A case study has been presented as proof of concept of supporting task analysis with Ambulatory Assessment tools. The study showed high compliance rates for respondents, and showed how valuable insights can be obtained that would be impossible or impractical to obtain otherwise. We have proposed a method for carrying out this task, which involves the articulation of a task model hypothesis, the construction of a context sensitive questionnaire, and the immediate benefits to facilitate contextual observations/ interviews. Future research should provide more detailed methodological guidance for task analysis procedures, ways of relating question-asking protocols directly to task modelling notations. Further case studies are needed to warrant the generalizability of the approach for addressing different work contexts and different needs of the analyst.

ACKNOWLEDGMENTS

The authors warmly thank the anonymous reviewers for their constructive comments on this manuscript. Suzanne Kieffer gratefully acknowledges the support of Wallonie-Bruxelles International (WBI) Excellence Grant which has made this collaboration possible. This research was partially supported by the project RW DGO6 Virtuoso-Plan Marshall convention 6608 funded by Région Wallonne.

REFERENCES

- Alliger, George M., and Kevin J. Williams. Using signal-contingent experience sampling methodology to study work in the field: A discussion and illustration examining task perceptions and mood. *Personnel Psychology* 46, 3 (1993), 525–549.
- Annett, J. Hierarchical Task Analysis. In *The Handbook of Task Analysis for Human-Computer Interaction*, (2004), 67–82.
- Batalas, N. and Markopoulos, P. Considerations for computerized in situ data collection platforms. In *Proc. EICS 2012, ACM* (2012), 231–236.
- Bevan, N. International standards for HCI and usability. *International Journal of Human-Computer Studies* 55, (2001), 1–18.
- Bias, R and Mayhew, D. *Cost-justifying Usability. An Update for the Internet Age*. Morgan Kaufmann (2005).
- Bodart, F., Hennebert, A.-M., Leheureux, J.-M., Provot, I., Vanderdonckt, J., Zucchini, G., Key Activities for a Development Methodology of Interactive Applications, Chapter 7, in Benyon, D., Palanque, Ph. (Eds.), “Critical Issues in User Interface Systems Engineering”, Springer-Verlag, (1995), 109–134.
- Carter, S. and Mankoff, J. When participants do the capturing: The Role of Media in Diary Studies. In *Proc. CHI 2005*, (2005), 899–908.
- Cooper, A., Reimann, R., and Cronin, D. *About Face 3: The essentials of interaction design*, (2007).
- Dallin, Gary W., *Control and Treatment of Hot-Dip Galvanized Surfaces*. Galvanizers Association Proceedings, (2005).
- Florins, M., Montero, F., Vanderdonckt, J., Michotte, B., Splitting Rules for Graceful Degradation of User Interfaces, *Proc. AVI’2006*, (2006), 59–66.
- Hudson, W. Card sorting. *The Encyclopedia of Human-Computer Interaction*, 2nd Ed., (2013).
- Johnson, H. and Johnson, P. Task knowledge structures: Psychological basis and integration into system design. *Acta Psychologica* 78, (1991), 3–26.
- Johnson, P., and Johnson. H. Knowledge analysis of tasks: task analysis and specification for human-computer systems. In *engineering the human-computer interface* (1991), 119–144.
- Jordan, B. and Henderson, A. Interaction Analysis: Foundations and Practice. *Journal of the Learning Sciences* 4 (1995), 39–103.
- Khan, V., Markopoulos, P., Eggen, B., IJsselstein, W., and de Ruyter, B. Reconexp: a way to reduce the data loss of the experiencing sampling method. In *Proc. MobileHCI 2008*, (2008), 471–476.
- Larson, R. and Csikszentmihalyi, M. The Experience Sampling Method. *New Directions for Methodology of Social & Behavioral Science*, 15, (1983), 41–56.
- Markopoulos, P., Pycock, J., Wilson, S., and Johnson, P. Adept-a task based design environment. In *Proc. HICSS 25*, (1992).
- Mayhew, D.J. *The usability engineering lifecycle: a practitioner’s handbook for user interface design*, (1999).
- Montero, F., López-Jaquero, V., Vanderdonckt, J., Gonzalez, P., Lozano, M.D., Limbourg, Q., Solving the Mapping Problem in User Interface Design by Seamless Integration in IdealXML, *Proc. of DSV-IS’2005*, (2005), LNCS Vol. 3941, 161–172.
- Paternò, F., Mancini, C., and Meniconi, S. ConcurTaskTrees: A diagrammatic notation for specifying task models. In *Proc. INTERACT 1997*, (1997), 362–369.
- Pinelle, D., Gutwin, C., and Greenberg, S. Task analysis for groupware usability evaluation: Modeling shared-workspace tasks with the mechanics of collaboration. *Transactions on Computer-Human Interaction*, 10, (2003), 281–311.
- Rieman, J. The Diary Study: A Workplace-Oriented Research Tool to Guide Laboratory Efforts. In *Proc. CHI 1993, ACM Press* (1993), 321–326.
- Trull, T.J., and Ebner-Priemer, U. Ambulatory Assessment. *Annual Review of Clinical Psychology*, Vol 9 (2013), 151–176.
- Vanderdonckt, J. Model-Driven Engineering of User Interfaces: Promises, Successes, Failures, and Challenges. In *Proc. 5th Annual Romanian Conference on Human-Computer Interaction*, (2008), 1–10.
- Wixon, D., Holtzblatt, K., and Knox, S. Contextual design: an emergent view of system design. In *Proc. CHI 1990, ACM Press* (1990), 329–336.