Domain Expert vs. Layman: Exploring the Effect of Subject Selection in User Studies for Industrial Wearable Computing Applications

Hendrik Witt
TZI Wearable Computing Lab.
University of Bremen
Am Fallturm 1, 28359 Bremen, Germany
hwitt@tzi.de

Ernesto Morales Kluge
Bremen Institute of Industrial Technology and Applied Work Science (BIBA)
Hochschulring 20 , 28359 Bremen, Germany
mer@biba.uni-bremen.de

ABSTRACT

We report on the results of a user study conducted to evaluate three different wearable input devices for menu-selection tasks in aircraft maintenance. Participating subjects were recruited from real maintenance workers (domain experts) and "typical" subjects of scientific studies (layman).

An unfamiliar data glove device using gestures was found to outperform natural speech interaction. Besides suffering from handling problems, speech interaction was found inappropriate for domain experts due to lacks in education; laymen did not suffer from this. Further, our evaluation suggests the necessity to include real end-users in studies for industrial applications to get results able to uncover application domain specific constraints. Study outcomes significantly differed between domain experts and laymen; suggesting to take social and educational backgrounds of end-users into account when designing user studies that should impact professional wearable computing applications.

1. INTRODUCTION

Years ago social and business contacts were mainly dominated by local relationships. People shared common languages and cultures. Nowadays, our local communities have become global; bringing together people with different cultures and languages spoken. With this, new challenges in human-computer interaction (HCI) arise with regards to the proper integration of both technological improvements and social aspects into usable applications.

The industry for producing and maintaining aircrafts is one domain that is affected by such social issues. With its widespread facilities all around the world, companies face cultural and language problems. Although being dominated by English documents, reports, etc., local languages and cultures still do exist and may impact application design. To support aircraft maintainers during their daily mobile work, wearable computing is deemed to be a promising technology (see e.g. [9]). Wearable computers can be unobtrusively worn on body to support a user during work. Unlike to stationary computing where users mainly concentrate on one task to be performed with the computer, wearable computing typically expects users to accomplish two different tasks [13]. A primary task involves real world physical actions, while the secondary task is often dedicated to interaction with a wearable computer. These unique characteristics, however, call for new interaction concepts and user interfaces. Findings from stationary HCI can often not be applied directly to interface design for wearable computers [13]. Thus, new user studies examining particular properties of HCI for wearable computers are needed. The wearIT@work project [15] was arranged to investigate these issues with an user-centered approach. It is supposed to empower mobile workers with real wearable computing applications. Here, questions particularly arose with regards to the impact the real end users involved in the application development process will have on experiments conducted to investigate how wearable interfaces and applications have to be designed to meet end user requirements.

In this paper, we focus on exploring the effect of subject selection for user studies tailored to examine interface interaction aspects for industrial wearable applications. Do experiment outcomes and drawn conclusions change when real end users (domain experts) are selected as subjects instead of the "typical" subjects that participate in scientific user studies? To approach this question, we report on an experiment that evaluated three different wearable input devices for navigation in wearable user interfaces of menu-selection style.

The remainder of this paper is structured as follows: Section 2 and 3 provide an overview of the field of research and the environment where the research presented was carried out. In section 4 the design of the experiment is presented which targets the pradigm of the primary and secondary task in wearable computing. The user study is presented in section 5 and contains a classification of the involved subjects as well as the description of the used apparatus. Section 6 presents and discusses the results of the study. Finally, section 7 concludes the paper and points out some future work.

2. RELATED WORK

Unlike with standard desktop applications that can be operated by keyboards and mouse devices, wearable computer call for different interaction devices that allow to be operated on the move. The Twiddler [7] was proposed as a keyboard for wearable text input and utilizes a chording principle. Speech interaction is a very unobtrusive and promising wearable interaction method [12]. One of its mayor advantages is, that it allows the user to enter text or application controls without using hands, i.e. it allows real hands-free operation [14]. Gesture interaction has been successfully applied in different wearable applications as well. In [9] a data glove devices was used to control an aircraft maintenance application. Several data gloves using different hardware components have been introduced so far [17, 3, 1, 11] that all allow the control of wearable applications.

Although many different interaction devices are already available for wearable computers, evaluations or comparisons of these devices are rare. In [7] Lions et al. compared the Twiddler with other mobile computing devices for text input. In [18], three wearable pointing devices were evaluated for selection tasks. What is common to these but also other user studies conducted for the scope of wearable computing in the scientific and university area is the subject selection procedure applied (see e.g. [5, 6, 10, 4, 2]). Usually, students are recruited from the local research groups or university for participation in experiments. These subjects, however, are at least to some extend biased by their educational background or interests. Regarding the outcome of experiments and their transferability to specific application domains the question arises how valid results become with these "typical" subjects; do they really behave similar to professionals?

3. BACKGROUND

One core driver of the European funded wearIT@work project is the "people first" vision explicated in the European ambient intelligence programm. To follow this vision, project outcomes of wearIT@work will be tested and proven by realistic pilot applications in a number of practical and relevant application domains where wearable computing has the potential to enhance existing processes. Thus, our research concentrates on the development of applications that follow a user-centered approach with focus on the mobile worker and his/her needs. For this, project partners cover various multidisciplinary backgrounds involving experts from computer science, engineering science, cognitive science, ergonomics, design, psychology, and sociology.

The research results presented in this paper impact the domain of aircraft maintenance which is one of four industrial scenarios of the project. For the scope of this paper, we do not focus on the envisioned pilot application for the aircraft maintenance scenario (see [8]), but on the results from a user study exploring the effects of subject selection on an experiment testing different interaction devices for application control of the envisioned application. The main driver for our experiment was the missing information about how aircraft maintainers feel and behave when using specific wearable user interface that are usually controlled by novel interaction devices beyond those known from desktop systems.

To examine these questions we started with interviews and workplace studies at an aircraft maintenance service facility to identify maintenance procedures suitable for a realistic experiment to evaluate different wearable input devices. Due to the involvement of domain experts in the experiment a set of real maintenance procedures was needed to motivate maintainers for participation. The main requirement of domain experts expressed during interviews regarding a wearable computer use was that is has to support and ease maintenance tasks and should be easy to operate.

To derive meaningful and more robust results from our planed study we had to enlarge the number of users participating in our study because access to domain experts was restricted to 10 users only. Therefore, we added additional users representing "typical" participants of research studies.

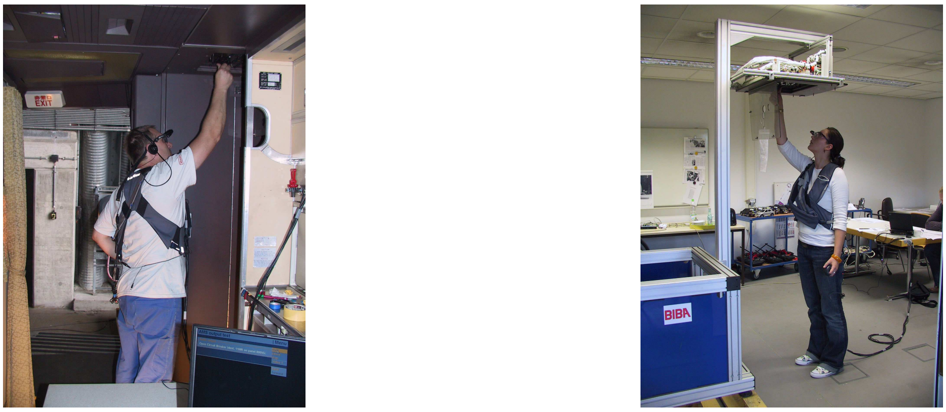
4. EXPERIMENT

The experiment addresses how different (partly novel) wearable input devices affect a person's ability to control a simple wearable user interface of menu-selection style that guides through maintenance procedures. The scenario involves the user performing a primary maintenance task in the real world, while a wearable computer guides the user through an aircraft maintenance procedure and requires the user to navigate within an application. By observing the user's performance in the primary task and the secondary computer task, conclusions can be drawn on what input devices are appropriate to use. Deeper analysis of the navigation accuracy also reveal strengths and weaknesses of input devices. To measure user performance in both tasks, an experimental model is needed. This section describes each task and how they are combined in the experiment.

4.1 Primary Task

As mentioned above, the design of the primary task was tailored to ensure maintainers' motivation for the experiment. The task used for this purposes was taken out of the Aircraft Maintenance Manual (AMM) of an Airbus A300; both, aircraft and manual, were available for the experiment. For keeping the experiment manageable and the results reproducible, a subtask of the passenger seat removal procedure was taken. The selected procedure fitted the requirement of being easy to learn for enabling also laymen and non cabin experts (although being maintenance operators) to perform the experiment. The chosen subtask, concerning the 800VU Panel (figure 2(a)) allows to simulate a complex working situation within the cabin of an Airbus A300 as well as under laboratory conditions. Figure 1 illustrates both primary tasks; one being performed in a real aircraft, the other under laboratory conditions.

The whole seat removal procedure comprises mechanical work steps as well as electrical work steps – currently for each domain, one expert is needed. Having in mind that the introduction of wearable systems should have a real impact on work processes, e.g., reducing the number of experts for a specific task, the electrical work steps fitted best for the experiment. They offered the needed complexity and the danger of messing up buttons, especially for cabin experts which currently need one expert (electrician) for this simple task.



(a) Within Aircraft (b) Within Laboratory

Figure 1: Subjects performing the primary task.

The electrical subtask covers the identification of the proper buttons on the circuit breaker panel to break the circuits of the systems that are connected to the seat. In a following step the electrical connections of the seat can be removed. Then, the mechanical task of removing the seat goes on.

4.2 Secondary Task

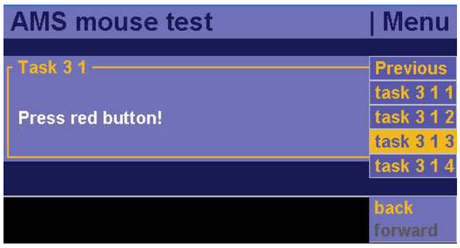
The secondary task consists of a menu navigation task presented in the users' head-mounted display (HMD). To avoid bias of study data that is caused by an inappropriate user interface design used for our navigation task, the WUI-Toolkit [16] developed in the wearIT@work project was used to generate a HMD optimized interface. The WUI-Toolkit offers a model-driven approach where application developers only have to specify an abstract model of the envisioned user interface; the toolkit itself then automatically generates a corresponding interface that is optimized for wearable computing applications.

An example of the secondary task is shown in figure 2(b). On the left side of the interface an instruction is shown that tells the user on which button on the circuit breaker panel (figure 2(a)) she is requested to press. The right part of the interface provides the actual menu for the navigation task. Dependent on the current test condition, the length of the presented menu varies between 4 to 7 menu items. To avoid confusion, menu items are labelled rather abstract instead of using application domain specific technical terms. Although the latter would be more realistic, because of their semantical meaning, it cannot be guaranteed that this is true for all users regardless of domain specific knowledge or jargon. Thus, labels are always structured in the same way; starting with the word "task" followed by a sequence of digits encoding the menu hierarchy. A yellow cursor provides feedback on the currently active menu item. If an item is selected, the user will be guided to the next level in the menu structure. Once a new level is presented, the cursors is placed by default on the first item in the menu.

5. USER STUDY

A total of 30 subjects were selected for participation – 21 males and 9 females aged between 18-52 years (mean 33.5 years). We selected 10 maintenance operators from an aircraft maintenance facility and 20 subjects from the local university. The three user groups represent different levels of expertise in aircraft maintenance and computer use and are structured as follows:

(a) Circuit Breaker Panel



(b) Navigation Task

Figure 2: Primary and secondary task.

1. **Domain Experts**
   (Professionals in aircraft maintenance), Group 1 (G1). The first group, consisting of 10 maintenance operators, was composed of professionals dealing with aircraft maintenance in varying domains (structure, cabin, engine, etc.).

2. **Laymen**
   (Subjects with background in engineering), Group 2 (G2). The 10 subjects of the second group were recruited at the campus of the local university in the fac-

ulties of mechanical engineering and information science; though people with an engineering background.

3. **Laymen II**
   (Subjects without background in engineering), Group 3 (G3). A third group, also recruited at the campus, consisting of 10 people without an engineering background (sociology, economy, cultural science, history) was selected to take part in the experiment.

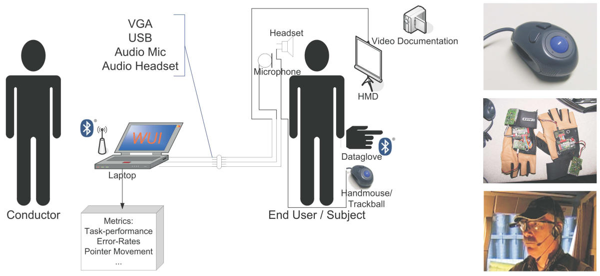
The study uses a within subjects design with the input device and navigation task length as the two independent variables, meaning that all subjects will test every combination of input device and navigation task length. A single test session consists of one practice round where the subject gets to practice the input device, followed by one experimental round during which data is collected for analysis. The time to complete a navigation task naturally varies depending on how quick the subject is, but on average pilot studies indicated it takes around 60 for the short and 240 seconds for the long navigation task for each device. Designing navigation tasks of different lengths allows to observe learning effects and fatigue caused by cognitive overload or physical constraints. A short task requires 4 navigation steps, a long 14 steps. The total time required for a session is around 20-35 minutes.

Instructions on the particular sequence to navigate in the menu are given orally by the experiment instructor immediately after the user has pressed a button on the circuit breaker panel. Because instructions were given in German and have to be translated by the user to a corresponding menu entry in English, the navigation task requires always some mental effort to be answered. A delay that might be caused by reaction time of the experiment instructor after a user has pressed a button on the panel, is deemed to be negligible, because the instructor was the same for the entire experiment and thus should have approximately equal reaction times for all users. To reconstruct potential faults (e.g. interruptions or delays) during the experiment the display of the laptop (the same UI the subject see on the HMD) and the subject while acting, were filmed with a video camera.

Each test consisted of a navigation task that included the following steps: (1) The conductor tells the subject which task to chose (e.g. "task 3.1.3") in the menu, (2) The subject then has to navigate to the corresponding menu item and has to "click" on it, (3) immediately the instruction appeared on the HMD (e.g. "press red button") (4) The subjects search the button on the panel and press it. The navigation includes also menu entries with no instruction to avoid that subjects get familiar with the navigation structure.

5.1 Apparatus

The original test environment in the aircraft was located in the front of the aircraft between the catering area and the cockpit. The circuit breaker panel (800VU) that was used in our experiment was mounted on the ceiling at a height of 2.05m. The test setup used in the aircraft had to be rebuild at the laboratory to get comparable results for the two reference groups (G2, G3). In figure 1 both test setups are illustrated. In the laboratory tests the 800VU circuit breaker panel was mounted on a stand at the original height of 2.05m. This guaranteed similar environmental conditions for the experiments.



(a) Hardware Test Set-Up. (b) Devices.

Figure 3: Experiment Test Set-Up.

In the experiment setup (see figure 3), the user is not wearing a wearable computer per se, as the HMD and the devices are connected via a 5m cable loom, containing USB (for the handmouse), VGA (for the HMD) and audio send and receive for the headset, to a stationary computer running the experiment. Although the wires and cabling for the HMD and the devices are still coupled to the user to avoid tangling, this should not influence outcomes compared to a setup with a truly wearable computer. In particular, we used a special textile vest the users had to wear during the experiment that was designed and tailored to unobtrusively carry a wearable computer, as well as all needed cabling for a HMD without affecting the wearers mobility. For having an even more realistic situation we put a OQO micro computer model 01[1] in the vest to simulate also the weight a wearable computer equipment would have outside the laboratory environment. The menu navigation tasks are presented in a non-transparent SV-6 Monocular HMD from MicroOptical[2]. The input devices for the experiment cover a pointing device (wearable trackball, handmouse), a gesture recognizing device (data glove) and speech recognition (see 3(b)).

6. RESULTS

After all data had been collected in the user study, the data was analyzed to study which effect the different input devices had on user performance and navigation accuracy. Because we were particularly interested in differences between user groups to see if overall results represent our domain experts, results were additionally analyzed within and compared between different user groups. For our analysis, the following metrics were used:

- **Time**: The time required by subjects to complete the navigation task from start to end, i.e. task completion time (TCT).
- **Accuracy**: The accuracy achieved by subjects during navigation, i.e. how often they selected not required menu items by accident.

[1] http://www.oqo.com
[2] http://www.microopticalcorp.com/Products/vga.html#SV6

Finally, a comparison of quantitative results and qualitative data gathered from questionnaires an post-hoc interviews was analyzed to draw additional conclusions and to put results in context.

6.1 Quantitative Analysis

To investigate differences among input devices used and differences between domain experts and laymen, we compared metrics separately for all subjects, domain experts, and laymen. The latter group was composed from the laymen groups G2 and G3 participating in the experiment. For the sake of simplicity, we henceforth refer to them as laymen.

6.1.1 Time

Our examination of task performance started with a repeated measures ANOVA to see whether there existed any significant differences among the input devices used. The ANOVA on short navigation data could only uncover a significant difference ($F(2, 57) = 3.933; p < 0.05$) in the laymen group. Post-hoc pairwise t-tests showed that laymen were significantly faster ($p < 0.05$) with a mouse device than with a data glove; supporting our implicit hypothesis that a mouse device is likely to outperform novel interaction devices in our semi-mobile experiment setup due to familiarity and missing anxiety during use. This was particularly true for laymen who's daily working background was characterized by frequent computer use in contrast to our domain experts that typically work outdoors without any computer support. Although our study could not uncover a significant difference, figure 4(c) indicates that laymen were also faster with speech interaction than with a glove for short interaction sequences.

Unlike average task performance of all subjects and laymen group (see figure 4(a) and 4(c)) that both performed on average worst with a glove device in short navigation, domain experts did not. Domain experts performed equally well with glove and mouse, but were slower when using speech. This suggests that differences in performance between domain experts and laymen do exist and can not necessarily be derived from performance results of all subjects; Laymen were 14.98 sec. slower with a glove than with the mouse. Due to differences in daily working procedures, domain experts apparently have no clear a priori preference for using a certain device. Instead, they started experiencing and learning each device at almost the same level in our study. The advantage of laymen that use mouse interaction during their daily computer work was not given for domain experts and was deemed to cause these performance differences. Although domain experts needed nearly always more time to complete short navigation tasks than laymen regardless of the device used, they did not suffer from previous experiences that can impair adoption or acceptance of novel interaction devices typically used in wearable computing applications such as data gloves. Indeed, several domain experts stated in post-hoc questionnaires and interviews that even though mouse was very easy to use, they are confident that they would have increased their performance with a glove very quickly if a bit more learning time had been allocated during the study.

Even though short navigation task analysis was already able to uncover significant differences and preferences for an input devices both, between groups and overall, an examination of user performance for long interaction periods allowed us to derive more conclusions.

For long navigation tasks, an ANOVA uncovered strong significant differences ($F(2, 87) = 15.178; p < 0.001$) on all subjects without group membership considerations. To investigate these differences in more detail, paired samples t-tests were performed comparing the different devices. Again, this was done for the three groups of interest. The results are shown in table 1. T-test results for all subjects (table 1(a)) showed significant performance differences that allow to rank devices very clearly. Similar to short interaction, best performance was achieved with a mouse device, followed by data glove and speech interaction. For long navigation tasks t-tests always showed stronger significant differences between all devices compared to short navigation tasks. Unlike to

Table 1: Pairwise t-tests on performance (time) for long navigation tasks.

(a) All Subjects

Time	**Mouse**	**Glove**	**Speech**
Mouse	-	**0.032**	<**0.001**
Glove	**0.032**	-	**0.014**
Speech	<**0.001**	**0.014**	-

(b) Domain Experts

Time	**Mouse**	**Glove**	**Speech**
Mouse	-	0.218	<**0.001**
Glove	0.218	-	<**0.001**
Speech	<**0.001**	<**0.001**	-

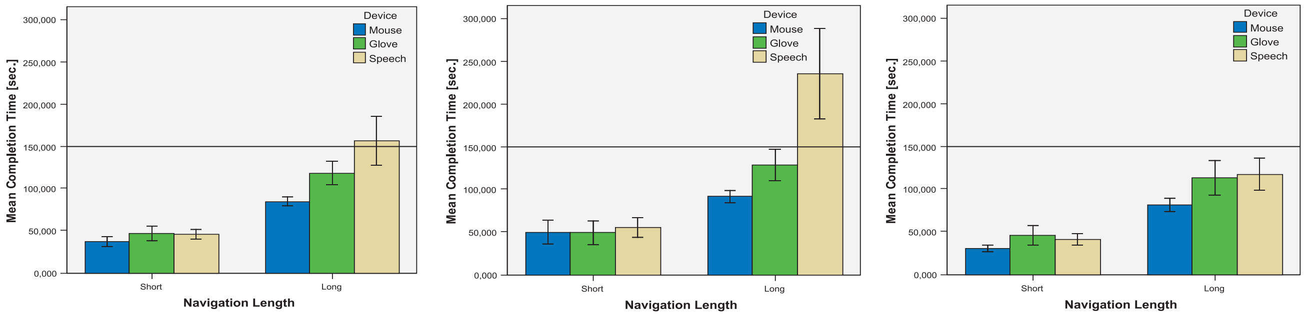
(c) Laymen

Time	**Mouse**	**Glove**	**Speech**
Mouse	-	**0.018**	**0.007**
Glove	**0.018**	-	1.000
Speech	**0.007**	1.000	-

short interaction where no clear advantage for either glove or speech interaction was found, glove interaction significantly ($p < 0.001$) outperformed speech interaction for long interaction sequences in all groups except for laymen (see table 1(c)). Once again this suggested that overall findings cannot necessarily be directly applied to specific user groups.

In general, speech interaction is prone to recognition errors of the speech engine that are likely to occur and impair performance more during longer interaction sequences than for shorter. Apparently, only domain experts suffered from this effect. Laymen did not perform significantly faster with a glove than with speech interaction. Actually, laymen perform on average 82.36% faster than domain experts (with respect to the performance achieved by domain experts).

An analysis of observation data collected throughout the study led to the conclusion that the discovered difference was at least partially due to varying English skills of subjects. Surprisingly, domain experts had huge problems with an appropriate English pronunciation of menu item labels when stating commands. This was completely unexpected, because pre interviews yielded that not only laymen but also all domain experts have sufficient English skills. Air-



(a) All Subjects (b) Domain Experts (c) Laymen

Figure 4: Average task performance of subjects separated in different groups.

craft manuals, reports, etc. are all in English. Although domain experts' self assessment during pre-interviews was to understand and speak English, our results show the opposite. Their skills impaired a reasonable use of speech interaction. Laymen did not have these problems.

Although the evaluation provided a clear ranking of devices for long navigation tasks, important differences between domain experts and laymen were found; indicating the necessity of including end-users in usability studies to get transferable results. For example, all subjects performed best with a mouse, but questionnaires data and post-hoc interviews indicated that several domain experts favored a glove device rather than mouse or speech interaction. Particularly, they judged a mouse as inappropriate for their daily mobile work on aircrafts due to its carrying constraints and the required mouse pointer control that they felt to cause fatigue on HMDs.

6.1.2 *Navigation Accuracy*

Navigation accuracy allows to determine the level of precision achievable with an interaction device. To analyze navigation accuracy, we computed the ratio of minimal required navigation steps to complete a navigation task with the actually needed number. For instance, if a user needed 6 steps to complete a task that actually required only 4 steps to be completed, the achieved accuracy was $4/6 = 0.67$, i.e. 67%.

An ANOVA performed on all subject's accuracy data showed strong significant differences for both navigation task length (short: $F(2, 87) = 20.659; p < 0.001$, long: $F(2, 87) = 22.565; p < 0.001$); indicating that differences in accuracy for both cases do exist. Post-hoc t-tests found strong significant differences ($p < 0.001$) between speech and the other devices regardless of task length. As expected, speech showed worst accuracy because of its operational differences to other devices. Unlike to mouse and glove interaction where two types of commands had to be issued to select a menu item (move cursor above an item and then select it), speech had lower "fault tolerance" and required only one command; a menu item was directly selected by just stating its label.

In line with figure 5(a) no significant difference could be observed in our study between mouse and glove interaction accuracy. With a mouse all subjects reached on average almost 100% accuracy for short and 99.17% for long navigation tasks, respectively. The glove provided very high accuracy at a level of approximately 95% for both lengths as well.

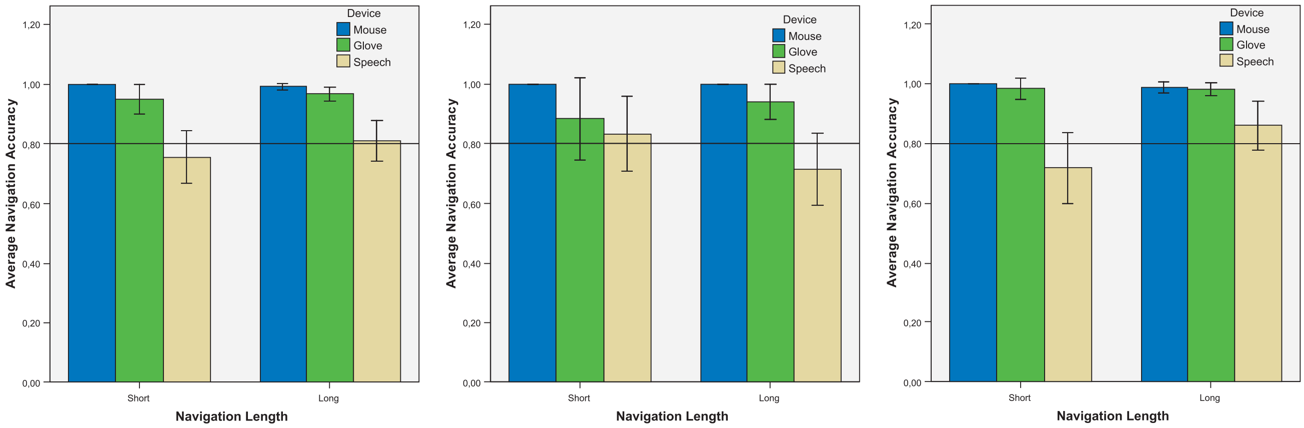
An anecdotal comparison of domain experts and laymen (see figures 5(b) and 5(c)) indicated no bigger differences between groups and overall accuracy. Mouse and glove were almost identical in accuracy regardless of navigation length. Unlike performance findings may suggest, our analysis could not uncover a significant statistical difference.

A detailed examination of speech interaction, however, discovered a surprising difference in average speech accuracy between domain experts and laymen. Accuracy for domain experts decreased from 83.33% for short interactions to 71.53% for long. Accuracy results for laymen behaved oppositional though. With 71.76% for short and 85.95% accuracy for long interactions, laymen were able to increase interaction accuracy. An interpretation, partly based on observations, yielded that this was once again due to different English skill levels. Laymen with higher English skills had already a personal speaking style that sometimes differed from recognizer expectations and consequently caused errors that impaired accuracy. However, laymen were able to "understand" the recognizer while performing the navigation task and instantaneously adapted their speaking style and intonation. For this, a certain time and "sense for language" was needed and therefore affected long navigation sequences only. Domain experts did not have these skills. Thus, the increased accuracy for short speech interaction was mainly by chance. For longer interaction phases, recognition errors outweighed chance and impaired accuracy.

Overall, subjects achieved best accuracy with a mouse and worst accuracy with speech. Although a data glove was the most novel and typical wearable computing device tested, accuracy was comparable with the familiar mouse device. Accuracy differences between experts and laymen were not found as dramatic as with performance measures; making a glove overall the preferred wearable device for menu selection for domain experts as already partially indicated in collected qualitative data.

7. CONCLUSION

We presented an experiment to determine the most appropriate input device for menu selection tasks for a profes-



(a) All Subjects (b) Domain Experts (c) Laymen

Figure 5: Average navigation accuracy of subjects divided into group.

sional wearable computing application in aircraft maintenance. By including professional end users (domain experts) as one part of the test group, experiment results significantly changed.

For short navigation tasks a data glove device yielded similar results as a mouse device with respect to both performance and accuracy achieved by domain experts and laymen. Disregarding the base case mouse device, we found that for longer interaction sequences our laymen can either navigate with speech or with a data glove. Domain experts were in contrast evidently faster and more accurate with a data glove device than with speech interaction although speech was deemed superior due to its natural and intuitive style for non computer experts. Domain experts suffered from handling problems of the speech recognizer. Their lacks in education made natural speech interaction inappropriate for their daily work. The reason for this was found in social and cultural differences in the aircraft maintenance business. Although domain expert's mother tongue was German, maintenance instructions and reports are by default always in English. Domain experts arranged themselves over time with the English language driven work situation. However, this comprises only reading and understanding English, but not speaking. Laymen did not have these problems. They could speak English fairly well even though is was also not their mother tongue.

The nature of results suggests that unlike to today's best practice in subject selection for user studies where participants are typically recruited from the university area; recruiting at least some domain experts should be preferred. Domain experts may have entirely different perceptions of novel technologies such as wearable computing that once being omitted risk applicability of results even though being statistically valid. An experiment without the real end users, therefore, is likely not be able to uncover these inherent differences.

7.1 Future Work

Although our experiments yielded valuable results, future work will concentrate on further experiments with real end users to examine how wearable user interfaces including interaction devices have to be designed properly. Taking into account the findings of the presented experiment, end users will be kept actively involved in the design cycles of the aircraft maintenance pilot application. Here, we will investigate also the development of a new special purpose wearable computing input device that is even more appropriate and easier to use by domain experts than the tested data glove device.

Acknowledgement

This work has been partly funded by the European Commission through IST Project wearIT@work: Empowering the Mobile Worker with Wearable Computing (No. IP004216-2004). We thank all persons participating in the experiment; especially, the maintenance staff at AMS, Ahlhorn, Germany.

8. REFERENCES

[1] M. Boronowsky, T. Nicolai, C. Schlieder, and A. Schmidt. Winspect: A case study for wearable computing-supported inspection tasks. In *ISWC'01: 5th International Symposium on Wearable Computers*, pages 163–164. IEEE Computer Society, 8–9 October 2001.

[2] H. W. Bristow, C. Baber, J. Cross, J. F. Knight, and S. I. Woolley. Defining and evaluating context for wearable computing. *Int. J. Hum.-Comput. Stud.*, 60(5-6):798–819, 2004.

[3] W. Chou, T. Wang, and L. Hu. Design of data glove and arm type haptic interface. In *11th International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (HAPTICS 2003)*, pages 422–427, 2003.

[4] J. Clawson, K. Lyons, T. Starner, and E. Clarkson. The impacts of limited visual feedback on mobile text entry for the twiddler and mini-qwerty keyboards. In *Ninth IEEE International Symposium on Wearable Computers (ISWC'05)*, pages 170–177. IEEE, 2005.

[5] M. Drugge, H. Witt, P. Parnes, and K. Synnes. Using the hotwire to study interruptions in wearable

computing primary tasks. In *The Tenth International Symposium for Wearable Computers (ISWC)*, Montreux, Switzerland, October 11–14 2006. IEEE.

[6] S. Hwang, G. Lee, B. Jeong, W. Lee, and I. Cho. Feeltip: tactile input device for small wearable information appliances. In *CHI '05: CHI '05 extended abstracts on Human factors in computing systems*, pages 1475–1478, New York, NY, USA, 2005. ACM Press.

[7] K. Lyons, D. Plaisted, and T. Starner. Expert chording text entry on the twiddler one-handed keyboard. In *ISWC '04: Proceedings of the 8th International Symposium on Wearable Computers*, pages 94–101, Washington, DC, USA, 2004. IEEE.

[8] E. Morales Kluge and H. Witt. Developing applications for wearable computers: A process driven example. In *Proceedings of the 4th International Forum on Applied Wearable Computing (IFAWC)*, Tel Aviv, Israel, March 12-13 2007.

[9] T. Nicolai, T. Sindt, H. Kenn, J. Reimerders, and H. Witt:. Wearable computing for aircraft maintenance: Simplifying the user interface. In *Proceedings of the 3rd International Forum on Applied Wearable Computing (IFAWC)*, Bremen, Germany, March 15-16 2006.

[10] M. Nilsson, M. Drugge, U. Liljedahl, K. Synnes, and P. Parnes. A study on users' preference on interruption when using wearable computers and head mounted displays. In *PerCom*, pages 149–158, 2005.

[11] R. Rosenberg and M. Slater. The chording glove: a glove-based text input device. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, 29(2):186–191, 1999.

[12] N. Sawhney and C. Schmandt. Speaking and listening on the run: Design for wearable audio computing. In *ISWC '98: Proceedings of the 2nd IEEE International Symposium on Wearable Computers*, page 108, Washington, DC, USA, 1998. IEEE.

[13] T. Starner. The challenges of wearable computing: Part 2. *IEEE Micro*, 21(4):54–67, 2001.

[14] T. E. Starner. The role of speech input in wearable computing. *IEEE Pervasive Computing*, 1(3):89–93, 2002.

[15] wearIT@work. The wearit@work project: Empowering the mobile worker with wearable computing. IST Project (No. IP 004216-2004), 2004.

[16] H. Witt. A toolkit for context-aware wearable user interface development for wearable computers. In *ISWC'05: The 9th International Symposium on Wearable Computers. Doctoral Colloqium.* IEEE, 2005.

[17] H. Witt, R. Leibrandt, A. Kemnade, and H. Kenn. Scipio: A miniaturized building block for wearable interaction devices. In *Proceedings of the 3rd International Forum on Applied Wearable Computing (IFAWC)*, Bremen, Germany, March 15-16 2006.

[18] J. Zucco, B. H. Thomas, and K. Grimmer. Evaluation of three wearable computer pointing devices for selection tasks. In *9th International Symposium on Wearable Computers (ISWC 2005)*, pages 178–185, Osaka, Japan, 18-21 October 2005.