
Hendrik Witt
TZI, Wearable Computing Lab.
University of Bremen
D–28359 Bremen, Germany
hwitt@tzi.de

Michael Lawo
TZI, Wearable Computing Lab.
University of Bremen
D–28359 Bremen, Germany
mlawo@tzi.de

Mikael Drugge
Media Technology
Luleå University of Technology
SE–97187 Luleå, Sweden
mikael.drugge@ltu.se

ABSTRACT
We present the results of an empirical study investigating the effect of visual feedback and body postures on gesture interaction techniques in a dual task setup found, for example, in wearable computing. The conducted experiment uses a novel apparatus called “Hot Wire” that allows retaining the properties of wearable computing even in laboratory environments. Visual feedback was found to impair user performance and caused users to be caught in an attention demanding closed feedback loop once presented in a head-mounted display. Even though continuous feedback was not necessary for gesture interaction, users were unable to ignore it and remain focused on the primary task. The design of an alternative gesture recognition method using a body-centric frame of reference instead of a conventional static one to improve usability, is shown to have an opposed impact both on the performance and subjective perception of users. The presence of novel devices in gesture interaction, such as data gloves, is found to be a major source of erroneous gesture recognition due to unpredictable user behavior. Our detailed result discussion provides guidelines for designing better gesture interaction.

Keywords
Gestures, Mobile interaction, Data glove, HotWire, Wearable Computing.

Categories and Subject Descriptors
H.5.2 [Information Interfaces and Presentation]: User Interfaces—Input devices and strategies, Interaction styles

1. INTRODUCTION
Wearable computers have the potential to augment a user’s ability to perform certain tasks, by providing relevant means for assistance, support, and guidance for the task at hand. This paradigm of wearable computing differs from ordinary desktop computing in that the tasks performed are typically of nature, taking place in the real world and requiring the user to maintain focus on this primary task rather than on the computer. In turn, this imposes different requirements and constraints on the user interface of wearable computers, as they must be designed to be simple and reliable to operate in order not to impede the user. They are typically operated using novel input and output technology such as hand-held or head-mounted devices. If highly mobile control of such interfaces is needed, gestures are deemed to be a suitable interaction method, since they can be performed blindly based on kinesthetic feedback while at the same time being to some extent hands-free. A device often used to implement gesture interaction is a data glove using acceleration sensors to monitor the user’s hand gestures [9]. They track hand movements and translate these into input events sent to an arbitrary user interface. A problem with devices which utilize accelerometers with earth’s gravity as their fixed frame of reference in wearable applications, is that users risk losing their orientation when in the midst of performing a physical task in the real world, an airplane technician crawling inside a narrow and sloping section of the plane.

In this paper we explore interaction using hand gestures with optional feedback in regard to this problem. The research question is how to make interaction with a wearable computer more accurate in challenging environments, where the user is forced to perform highly mobile and physical tasks at the same time as operating the computer.

1.1 Hypotheses
Our hypotheses to be verified with a user study are:
H1. Providing continuous visual feedback will be detrimental to the user’s overall performance, and omitting such feedback will yield better performance in a typical wearable computing scenario.
H2. More accurate control can be achieved if hand gesture recognition uses a body-centric frame of reference for interpreting data, rather than a world-centric frame of reference.

2. MOTIVATION AND RELATED WORK
As humans naturally support verbal communication with gestures, they have been subject to research in human-computer interaction (HCI) for a long time. Pirhonen et al. [15] demonstrated that gestures are convenient for mobile communication as most people can perform them while being on the move. Compared to common stylus-based interaction techniques for mobile computing that require visual feedback and thus inhibit our movement, gestures do not. A very common result
of usability studies on interaction controlled by visual feedback is therefore that feedback interrupts users in their movement while interacting with the computer as the secondary task becomes the primary one. But what happens if the display that shows the feedback is even perceivable without moving the head such as in the case with head-mounted displays (HMDs) typically used in wearable computing?

In [2], Brewster et al. evaluated two novel interfaces using sound and gestures for wearable computers. They demonstrated that non-visual interfaces using audio feedback can be an effective alternative to visual-centric interfaces. Hence, suggesting that feedback improves interaction. Clawson et al. [3], however, showed in a user study that mobile text entry with mini-QWERTY keyboards is more sensitive to the absence of visual feedback with respect to typing speed than, e.g., when the Twiddler one-handed chording keyboard is used.

These examples demonstrate that in wearable computing feedback does not always improve or impair interaction, but seem to depend on both the type of feedback and interaction device used. In line with this, Marentakis et al. [10] also support this assumption with their work. They found that providing audio feedback for mobile interaction with a non-visual audio display decreased task performance rather than increased it while other metrics observed throughout their study yielded opposite results.

The effect of providing visual feedback for gestural interfaces is, however, deemed to provide proper user guidance. It is sometimes considered as necessity for user satisfaction because users do not sufficiently feel how they performed a gesture in space and may want to receive guidance in case the interaction failed. Kailio et al. [9] evaluated the use of visualizing entire hand motions of performed gestures as feedback for interaction. They argued that such visualization can be useful for gesture feedback but discovered also that users can get confused when obscure visualizations resulting from erroneous gesture recognition were presented. By contrast, it is known that users’ internal sensation of body postures and performed motions allows users to feel how they control an input device without the need of looking at it or using visual feedback. In particular, this has been found an important factor when users are involved in multiple tasks and therefore have to divide their attention [1, 6, 13].

In conclusion, this suggests that it is not yet entirely clear in advance what impact feedback provided for a certain interaction technique will have or how it interferes with other characteristics of the computing paradigm. For instance, one of the unique characteristics of wearable computing is its high degree of mobility involving different body postures of users being involved, e.g., in a primary manual maintenance task. Thus, we are interested in our experiment not only in the impact visual feedback might have on hand gesture interaction for wearable computing, but also if different body postures interfere with them or have an impact on the interaction technique itself.

2.1 Evaluating Gesture Interaction in Wearable Computing

Some work demonstrated the use of glove-based gestures to control user interfaces of wearable computers. For instance, Thomas and Piekarски [18] presented glove-based interaction techniques for an outdoor wearable augmented reality scenario. In [20] a data glove was used to operate vertical menus of an aircraft maintenance application by using two easy to learn hand rotation gestures. Gesture input using devices with a different form factor than data gloves were proposed also like, e.g., Gesture Pendant [16], Gesture Wrist [16], or FreeDigitier [12]. What is common to all these gesture devices and their used gesture recognition, is that sensor data is interpreted in a static manner, i.e. it is interpreted in a static world-centered frame of reference that neglects the highly mobile property commonly present in wearable computing environments. That is, an operation where the user is highly mobile and involved in a dual tasks situation with a primary physical task in the real world and a secondary computer task. That these properties can make a difference compared to stationary computer use, has also been argued by Witt and Drudge [19] who developed an apparatus called “HotWire” that allows evaluating aspects of wearable computing while retaining its properties even in a laboratory environment using a physical abstraction of manual tasks. In comparison to earlier studies [4, 14], Drudge et al. [5] found that by using the HotWire more differences can be uncovered. One of the differences reported was that different non-standard body postures of users had a negative impact on gesture interaction.

2.2 Different Frames of Reference in Gesture Recognition

The use of different frames of reference in user interaction has been recently demonstrated for cameras that automatically detect their orientation [8]. Here, the user can change orientation of the camera device but still maintain her body-centric frame of reference which allows for a much easier and convenient interpretation of the camera display.

Although not closely related to our work, up to now different frames of reference have been mostly studied for virtual reality and 3D user interfaces design. In [7], Hinkle et al. presented design issues for 3D user interfaces using different frames of reference and argued that “users may have trouble moving in a fixed, absolute coordinate frame”. To overcome this, Hinkle et al. proposed interaction techniques to be based upon motions relative to the user’s own body rather than absolute to a static world-centric frame of reference.

With regards to Hinkle et al.’s definition, figure 1 depicts these two alternatives of using different frames of reference for hand gesture recognition. Figure 1(a) depicts the commonly used absolute approach for gesture interaction where (angular) calculations are done by calculating the differences of the current sensor data to the fixed world-centric frame of reference defined by earth’s gravity. Opposite to this approach figure 1(b) shows the relative approach where sensor data is interpreted with respect to a body-centric frame of reference that is defined by the user herself. As indicated in both figures, calculations result in different values of angle α. In the absolute method the user has to compensate manually the angular difference β with respect to her own body posture, which adds additional cognitive load that might effect interaction accuracy. Applying the relative method instead, it compensates all body-movements of the user automatically in a way that all interaction is relative to her body posture. For example, for absolute a “thumb up” position is only reached once the thumb is turned back −45°, i.e. pointing toward the ceiling. For the relative method “thumb up” is reached when the thumb is in line with the user’s pitch, i.e. in the given case no adjustment is needed, because body and hand are tilted in the same angle. Note, that while the user maintains an upright posture, the absolute and relative methods are equal.

3. EXPERIMENT

The experiment attempts to establish whether body-relative gesture recognition (henceforth denoted as the relative method) will offer the user better control when using a data glove for interaction, compared to the current recognition (henceforth denoted as
the absolute method) where the user needs to maintain a sense of the hand’s orientation in relation to the real world and earth’s gravity. In connection with exploring the difference between the relative and absolute method, the experiment also evaluates the impact of providing visual feedback while performing the interaction gestures. To test this, a user is set to perform a physical task in the real world, while simultaneously being enforced to interact with a virtual task in the wearable computer. By comparing the user’s performance of both tasks when using relative as opposed to absolute recognition, some conclusions can be drawn about the feasibility and accuracy of the two recognition models. Furthermore, additional conclusions can be drawn on how visual feedback interferes with the user’s task performance, i.e. whether or not the presence of visual feedback in the relative or absolute gesture recognition could increase task performance.

3.1 Physical Task

The physical task needs to represent a typical situation encountered in different wearable computing scenarios. Examples of such tasks include, e.g., mechanical assembly of vehicles or routine inspection and audit of machine components. Often, such tasks require the close proximity of the user, typically occupy one or both of her hands, and requires a major part of her attention. Furthermore, the tasks tend to be mobile in that the user needs to move around as well as transport parts or components from one place to another. To represent this kind of task, the HotWire [19] experimental setup was chosen.

The HotWire apparatus simulates a primary task that fulfills the requirements of being mobile, physical, and attention demanding. It consists of a 3 meters long metallic wire bent in different shapes over which a metallic ring is meant to be passed by the user without touching the wire. When the user accidentally lets the ring touch the wire, that is seen as an error. Starting at one end of the wire, the completion time and number of errors can be recorded as the user progresses along the wire, until finally reaching the other end which signifies the completion of the task.

As the experiment is about the impact of body- versus world-relative positioning, our wire track is shaped such that it will enforce users to bend, stretch, and in general get into body postures that would make them more prone to lose orientation. Furthermore, the sections of the track are varied in terms of difficulty, exposing the user to easier as well as more tricky sections that require a higher degree of concentration on the physical task. Figure 2(a) shows the complete HotWire apparatus used in this experiment.

3.2 Virtual Task

The virtual task needs to be able to receive the input events relevant to the experiment. In our case that are the gestures capable of
The matching task is presented for the user in a HMD, as such output devices are commonly used in various wearable computing applications (2(b)). Here, three figures are shown of random shapes and colors, and the user must match the figure on top with either the left or the right figure at the bottom of the display. A text instructs the user to match either by color or by shape, and as there are 3 possible shapes and 6 colors, this makes the task always require some mental effort to answer correctly. To increase the cognitive workload of the user on more levels than just shape and color matching, a second kind of matching task was added in form of a calculus exercise (see figure 3(b)). The calculus task presents a mathematical expression of type

\[ X \text{<operator>} Y, \text{<operator>} := + | - | * | / \]

Below this expression one correct answer is given and one erroneous answer assigned randomly to left and right. The user answers by tilting the hand left or right, in the same manner as the aforementioned matching tasks are controlled. In the current experiment the expressions and answers were limited to integers only ranging from 1 to 9, for the sake of simplicity and ensuring mainly correct responses while still requiring enough mental effort of the subjects tested.

\[ \text{Match by shape} \]

(a) Figure matching

\[ \begin{array}{cc}
2 + 3 \\
5 & 7
\end{array} \]

(b) Calculus matching

Figure 3: Matching tasks representing the virtual task.

The matching tasks are considered abstract enough not to interfere cognitively with the physical task of the HotWire. It could be argued that the virtual task should be modelled as being more related to the physical task, e.g., providing guidance on how to best proceed over a difficult section of the wire. However, it was deemed more suitable to keep the task as simple and easy to comprehend as possible, in order to focus the study on pure performance metrics for the interaction method used, and avoid any bias caused by the users’ interpretation and handling of such guidance. Naturally, the findings from the experiment in terms of interaction style can then be directly applied to more specific virtual tasks in different application domains.

3.3 Sampling of Body Postures

The virtual task runs concurrently as the user performs the physical task so that matching tasks will be presented for the user at random time intervals. Each matching task will thereby interrupt the user and call for the user’s attention in handling it. As this will occur at random points on the track, a wide variety of body postures will be sampled at the time the matching task is presented and handled. As the user’s performance in both the physical and virtual task is monitored, this sampling makes it possible to determine how well the virtual task can be handled in various situations, using various input methods. In turn, this enables us to draw conclusions about the usability of using either the relative or absolute input method.

To increase the validity of the experiment, it is desirable to prevent the user from letting a queue of unanswered matching tasks build up. If a queue is allowed to build up which allows the user to answer them all in one batch one after the other, this could allow the user to attempt to “cheat” by using various strategies. Therefore, a time limit of 5 seconds was imposed causing the matching task to disappear and be counted as an error unless it is handled soon enough, to force the user to handle the matching tasks as soon as possible and provide correct answers. To avoid queues building up, all matching tasks were interspersed with 1 second plus a random period of time of up to 5 seconds. Combined, this means a new matching task will appear on average every 8.5 seconds while performing the physical task.

Another reason not to let queues of matching tasks build up is that this can otherwise cause a bias in the experiment.

3.4 The Data Glove Input Device

A pair of leather fitness gloves was used as the basis for our data glove prototypes. The sensor board including a battery pack was mounted on the back of the hand of each glove. A trigger button, providing some tactile and acoustic “click” feedback when being pressed, was positioned on the forefinger to be easily accessible. An accelerometer gathered motion data for the gesture algorithm while the trigger was used to actually issue a gesture.

To let interaction gestures correspond to the way matching tasks are presented, two gestures for selecting either the left or right object were chosen: By tilting the glove to either left or right and exceeding a predefined threshold angle (\(> 45^\circ\)), a gesture is activated. Once activated it can be finally issued to select an object by pressing the forefinger button. The two different gesture recognition algorithms tested in the experiment only differ in angular calculations. To determine whether a tilting angle is above the threshold the tilting angle is measured with respect to a “thumb up” position. In the absolute case “thumb up” is related to the static frame of reference that is defined by earth’s gravity. In the relative case, however, “thumb up” is measured with respect to the users body-centric frame of reference. That is, once the user tilts her head/body to one side the body-centered frame of reference detaches from earth’s gravity reference so that the “thumb up” position is reached once the thumb is in line with the current tilting angle of the users’ body (see figure 1(b)). To measure the required body-centric frame of reference, users have to wear a special headset with an acceleration sensor centered above the user’s head during the experiment.

3.5 Visual Feedback

To provide the user with real-time information on the gestures for selecting either the left or right object in the virtual task, visual
feedback could be provided in the HMD where the matching tasks are shown. The feedback consists of a continuously updated widget in which the data glove’s current tilting angle is visible, together with the frame of reference marking the neutral position where neither left nor right is selected. For the absolute method the frame of reference is the earth’s gravity vector, i.e. straight down regardless of how the user moves around (see figure 4(a)). For the relative method, the frame of reference follows the tilting angle of the user’s head (see figure 4(b)). The widget also shows two “activation areas” marked in red and green, that move along with the frame of reference. When the tilting angle (orange line) of the user’s hand reaches one of these areas, a colored square, corresponding to the color of the activation area, is shown in the center to let the user know an input event (left or right) can be triggered by pressing the button.

Figure 4: Visual feedback with reference angle fixed to earth’s gravity (left) or following the user’s head (right).

All in all, the visual feedback provided guides the user how to move to make a left or right selection. The feedback was deliberately not integrated into the actual matching task, e.g. through highlighting the left or right object, in order to serve as a “worst case” kind of feedback which may be more complex than needed for the situation. By testing the absolute and relative methods with and without this feedback, knowledge can be gained on what impact such feedback will have.

4. USER STUDY

A total of 22 subjects were selected for participation from students and staff aged between 23–43 years (mean 27.95). The study uses a within subjects design with the gesture recognition method and the inclusion or exclusion of visual feedback as the two independent variables. Eight treatments were used; one base case with only the physical HotWire task performed, three base cases with only the matching task performed (figure matching, calculus matching, and both combined at random), plus four experimental treatments consisting of the absolute and relative methods with and without feedback. Between each treatment the user had a chance to rest to reduce fatigue and prepare for the next treatment described verbally by one of the test leaders. To avoid bias and learning effects, the subjects were divided into counterbalanced groups where the order of the treatments differed. A single test session consists of one practice round of the physical and virtual task, followed by one experimental round during which data is collected for analysis. In the practice round, each treatment is performed once so that the user can learn the different treatment combinations. Although merely five runs over the HotWire are done in the practice round, this has shown to be enough for the subject to become sufficiently proficient to perform evenly for the remainder of the experiment. In the experimental round, each of the eight treatments is performed twice so as to yield more data for statistical purposes; in practice, this is equal to providing a twice as long track. The time to complete the physical task, as indicated by pilot studies, will take around 40–60 seconds for one single run over the wire. Overall, the total time required for a session is around 45 minutes.

4.1 Running the Experiment

The experiment was conducted in an isolated room at the local university, with only the subject and two test leaders present. Each subject was asked to fill in a form with basic demographical questions and estimated ability to perform motoric tasks in general. Then, the apparatus and the absolute and relative orientation methods and how they were operated were introduced. After this, the actual practice and experiment rounds followed. Finally, the subject was asked to fill in a form with questions regarding the experiment just performed. This form included questions about the user preference of the different orientation methods and the visual feedback provided.

5. RESULTS

After the user study, the collected data was analyzed to study the impact of visual feedback and how the different orientation methods affect user performance. The following metrics were used in the analysis:

- **Time:** The time required for the subject to complete the HotWire track from start to end.
- **Contacts:** The number of contacts the subject made between the ring and the wire.
- **Error rate:** The percentage of matching tasks the subject answered wrong.
- **Average age:** The average time from when a matching task was created until the subject answered it.

The first two metrics, time and contacts, refer to the subjects’ performance of the physical task. The last two metrics, error rate and average age, refer to how the virtual task was performed. The lower the value of a metric is, the better the subjects’ performance is in that respect. The graphs in figure 5 summarize the overall user performance by showing the averages of the metrics together with one standard error.

A Two-factor repeated measures ANOVA was performed over the four experimental treatments where the absolute and relative methods, with and without feedback, were tested. Significant differences ($F(1,84)=10.444$, $p<0.002$) were found for the average age metric in the performance of the virtual task. For the remaining three metrics this initial analysis did not reveal any significant differences. This is, however, partially because interaction effects between the physical and virtual task will show up most clearly in the average age metric, whereas for the other metrics they can be obscured and require a more in-depth analysis to reveal the differences. Therefore, to further investigate the data in more detail, paired samples t-tests were performed comparing each of the four treatments with each other. The results of the metrics where significant differences were found are shown in table 1. A Bonferroni corrected alpha value of $0.05/6=0.008$ was used to accommodate for multiple comparisons. In the forthcoming sections the metrics and differences will be examined in further detail.
5.1 Completion Time of the Physical Task

Related work has found that users tend to stop or slow down their motions when interacting with a device in mobile settings. Hence, the completion time will increase once the virtual task is introduced alongside the physical task, and this can also be seen in figure 5(a) showing the times for the base case of the physical task and the experimental treatments. Since matching tasks appear with the same frequency regardless of treatment, the overall slow down effect will be the same for all treatments. Thus, the individual differences in completion time among the four treatments lies either in the method or feedback used. Interpreting the results of the t-tests, we can see that the completion time of the physical task is negatively affected by the use of visual feedback, while nothing can be stated with certainty regarding whether the absolute or relative method is preferable. For the absolute method, feedback caused significantly (\(t=-2.75, p<0.008\)) longer time than absolute without feedback. This was also seen (\(t=-3.4, p<0.003\)) than the same with feedback again indicates that visual feedback is detrimental, supporting hypothesis H1. With feedback included, the absolute method was also significantly better (\(t=-2.63, p<0.008\)) than the relative method, providing one indication to disprove hypothesis H2 which states that the relative method would be better. Comparing the base case of match only (mixed calculus and figure matching) with the four treatments, an ANOVA showed a significant (\(F(4,105)=5.98, p<0.0003\)) difference in between the average age metric in the different treatments. Following t-tests revealed strong significant differences between the base case and the absolute (\(t=-3.28, p<0.002\)) and relative (\(t=-4.43, p<0.001\)) methods in the case where visual feedback was present. Since the base case does not involve the physical task, this factor alone can contribute to a higher average age overall, but does not explain the relative differences between the treatments themselves. These differences are thus caused either by the orientation method used or by the visual feedback. Because the two treatments where feedback is present exhibit a worse performance than the other two treatments without feedback, this suggests that the differences in average age between the treatments can be attributed to the inclusion of visual feedback. Again, this indicates that visual feedback is detrimental to performance, and cause a significantly higher time to handle matching tasks.

![Figure 5: Averages of user performance.](image)

5.2 Average age in the Virtual Task

By analyzing the average age of the matching exercises in the virtual task, we found that relative with feedback exhibited the very worst performance (see figure 5d), and that the remaining three treatments were all significantly better. The fact that relative without feedback was significantly better (\(t=-3.22, p<0.003\)) than the same with feedback again indicates that visual feedback is detrimental, supporting hypothesis H1. With feedback included, the absolute method was also significantly better (\(t=-2.63, p<0.008\)) than the relative method, providing one indication to disprove hypothesis H2 which states that the relative method would be better. Comparing the base case of match only (mixed calculus and figure matching) with the four treatments, an ANOVA showed a significant (\(F(4,105)=5.98, p<0.0003\)) difference in between the average age metric in the different treatments. Following t-tests revealed strong significant differences between the base case and the absolute (\(t=-3.28, p<0.002\)) and relative (\(t=-4.43, p<0.001\)) methods in the case where visual feedback was present. Since the base case does not involve the physical task, this factor alone can contribute to a higher average age overall, but does not explain the relative differences between the treatments themselves. These differences are thus caused either by the orientation method used or by the visual feedback. Because the two treatments where feedback is present exhibit a worse performance than the other two treatments without feedback, this suggests that the differences in average age between the treatments can be attributed to the inclusion of visual feedback. Again, this indicates that visual feedback is detrimental to performance, and cause a significantly higher time to handle matching tasks.

5.3 Error Rate and Contacts Metrics

Comparing the error rate on matching tasks between the four treatments shows that errors are consistently slightly higher for the relative method compared to the absolute method, regardless of whether feedback is used or not (see figure 5(c)). This differences

<table>
<thead>
<tr>
<th>Table 1: Pairwise t-tests of treatments.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>Abs w/o</td>
</tr>
<tr>
<td>Abs with</td>
</tr>
<tr>
<td>Rel w/o</td>
</tr>
<tr>
<td>Rel with</td>
</tr>
</tbody>
</table>

updated in real time and the user’s attention is caught in the virtual domain, thereby reducing the performance of the task located in the physical domain. The qualitative data from the post-experimental form reports that users could ignore the feedback, but apparently did not. It should be stressed that no user considered the feedback widget to be complex or difficult to interpret, on the contrary several users stated they fully understood and were able to use the information within it. Hence, the negative impact of visual feedback was not because of a bad implementation of the widget, but rather that it deprived them of the ability to focus their attention on the physical task. From a design perspective the feedback was also meant only to serve as guidance for the users when needed, and not something they would use all the time to perform the gestures. This indicates that visual feedback, even in a form that is not required for the task at hand, can cause problems when implemented.
are non significant though. Also no significant differences could be found for the number of HotWire contacts in the physical task. Even a more detailed examination of the metric could not uncover significant results. Thus, error rate and contact metrics were not able to provide additional inside for our hypothesis.

5.4 Effort of Hand Gestures

The four metrics discussed thus far are very relevant for the performance and accomplishment of a certain task, although they only tell about the user performance once an action has been made — not what events occurred or what interaction was being made before that action was performed. One aspect that a well designed user interface should help the user with, is to optimize user performance by avoiding needless or redundant interaction steps. For a user interface controlled via gestures, this means minimizing the movement of the user’s hands to perform an action. Ideally, the user would tilt the hand to left or right and push the trigger to issue the event and perform the desired action. If the user would for some reason tilt the hand back and forth a lot, this constitutes unwanted interaction steps that serve no purpose (unless there is feedback and a matching task to handle, in which case the user may do it simply for the purpose of testing or thinking). Thus, it is of interest to analyze how the user’s hand moved during the treatments, to test whether hypothesis H2 holds in its assumption that the hand will follow the user’s body posture to a higher degree in the relative method than in the absolute method.

The assumption is that the number of tilting gestures, henceforth called activations, will be reduced in the relative method as the user’s hand stays more coupled with the body’s frame of reference compared to the absolute method. Two t-tests performed comparing the absolute and relative methods first with feedback and then without feedback, for the case where no matching task was present, shows strong significant (t=2.79/-4.03, p<0.01) differences between the number of activations made. The relative method, however, exhibited a higher number of activations than the absolute method. This was contrary to our assumption; if the user’s hand would follow the body in the relative method the opposite situation would have occurred. As the user’s movement patterns are the same regardless of whether relative or absolute is used, the absolute method would exhibit a higher number of movements as the head motion alone causes activations\(^1\). Now, since relative exhibited a higher number of movements than absolute, it means the user’s head moves. Since the natural head motions as the user moves along the track can be assumed to be similar regardless of method used, this implies the user’s hand is held still and kept in the world relative domain rather than following the user’s natural body movements, which would disprove H2.

5.5 Accuracy of Hand Gestures

Delving further into the interaction patterns of hand movements, it is of interest to know how many activations (tilting the user’s hand above the threshold) are needed to handle a matching task. Ideally, exactly one activation is needed — tilting to the correct side directly when the task appears. To analyze this, we only consider the periods of time during which a matching task was present in the HMD, and count the number of all activations needed from the appearance of the task until it was responded to. In this case, it is the number of activations divided by the total number of matching tasks that is the relevant metric to compare among subjects. Note, that we do not divide by time, as that would give a performance related metric that can vary greatly between subjects, and which is not of interest when studying user behaviour of the data glove as a concept. The case where no interaction occurred and the matching task timed out was ignored, and only tasks actually answered (correctly or not) are counted.

Performing pairwise t-tests between the four experimental treatments did however not reveal any differences, as the users appeared to have answered each matching task in a similar manner regardless of treatment. In particular, the relative method was not worse in terms of accuracy when disregarding the response time, meaning that it is not necessarily less accurate compared to the absolute method.

6. DISCUSSION

6.1 The Problem of Immersive Feedback

A number of statistically significant differences were found when comparing treatments with and without feedback, indicating that the presence of the visual feedback widget reduced user performance with respect to many metrics. The most obvious metrics affected were those related to performance; the completion time of the HotWire physical task, and the average age of a matching task until it was handled. Although the visual feedback widget was not integrated into the matching tasks, but rather shown below on the same HMD, the user could attempt to ignore it. Despite this, the presence of the feedback widget still proved to be detrimental for the user’s performance of the physical task. According to the qualitative data from the forms, many users indicated that in general they did not like the use of visual feedback, and that they could and did ignore the feedback as it was not needed. The latter is contradicted by the quantitative data which clearly shows that the feedback still had a negative impact on their performance, whether they were aware of it or not. Of interest is also that users considered feedback to have a more negative impact on the physical rather than the virtual task, making the former harder to perform even though the feedback has no relation to it. The probable cause for this difference in opinion is that when focusing on the physical task, the visual feedback would still be present and disrupt the user’s attention.

The feedback in its current incarnation, where it is continuously updated yet where every frame can be understood separately, constitutes a closed feedback loop that requires the user’s constant attention. Therefore, we argue that providing this kind of closed-loop visual feedback in wearable computing scenarios can cause problems with the physical task. Continuous visual feedback for gestures should not be used if the primary task is attention demanding. This opinion has been known anecdotally in the wearable computing community [17]; our study validates this and also points out that merely the presence of visual feedback can be detrimental, even in cases where the feedback is not necessary to successfully carry out the virtual task. The users were not able to ignore the visual feedback when performing the primary task, even though they reportedly tried and considered themselves to manage it successfully. Alternative methods such as tactile or audio based feedback may be preferable, as these neither require the user’s visual attention nor stimulate her visual system to the same degree. However, audio feedback also exhibits problems as reported in [10], indicating that feedback is still a problematic issue to implement correctly in mobile settings.

\(^1\)Note that for the relative method, the angular difference between the hand and head’s frame of reference is used, and an activation can thus be caused either by tilting the hand or tilting the head.
6.2 The Novelty of Interaction Techniques

No evidence has been found that our implementation of the relative body-centric orientation method has any advantage over the absolute method that uses a commonly known world-centric frame of reference. One of the reasons for this, is that the fixation of the user’s hand in the real world’s frame of reference was not the natural behavior we expected when hypothesis H2 was formulated. Post-experimental video and photo analysis of a number of subjects confirm this finding: the vast majority of users held the hand wearing the data glove unnaturally still, with their hand and/or arm straight out roughly aligned to the horizontal plane, in an unrelaxed and “robot-like” manner. This behavior will indeed cause a much higher number of activations to occur with the relative method used, as the user’s head moves along with the track whereas the hand is held and maintained in a static position. This was likely because of the novelty of the device, which made users use it in a manner they considered appropriate yet which was still unnatural. For the absolute method, this behavior makes for an ideal performance as the user’s hand is always kept aligned with the real world, even though this would not be a natural movement pattern for long-term use. This finding is supported by the analysis of the objective data collected, as well as through video and photos captured during the subjects’ performance of the experiment. In the qualitative data, subjects consistently ranked absolute as being better than relative in terms of overall liking and ease of use. On the question whether the data glove was difficult or easy to use, the users did however find it easy or very easy to use on average, indicating that they did not consider their “unnatural” way of interacting with it to cause any problems. Despite this, they still indicated a strong dislike for the relative method, likely because they did not use the interaction device as it was designed to be used and therefore found it more troublesome to use. As they could not see the actual difference between the algorithms being used, their unfamiliarity with the relative method was what caused this negative feeling towards it.

7. CONCLUSIONS

We have presented the results of an empirical study investigating the effect of visual feedback and body postures on gesture interaction techniques in a dual task setup. Visual feedback in a head-mounted display was found to impair user performance, the reason being that its presence caused the user’s attention to be caught in a closed feedback loop. Even though continuous feedback was not necessary for successfully carrying out the gesture interaction, users were unable to ignore it and remained focused on the primary task. Our recommendation is to use visual feedback with care and avoid continuous and closed feedback loops in this kind of settings, and that alternative means for feedback should instead be investigated further. Furthermore, employing an alternative gesture recognition method using a body-centric frame of reference instead of a conventional world-centric one to improve usability, was shown to have a negative impact both on the performance and subjective perception of users. The primary reason for this was not technical in nature, but rather that users behaved both unpredictable and unnatural when using a novel interaction device which they may not have had experience with beforehand. We note that short term user studies may not accurately reflect long-term natural usage, and recommend interaction designers to consider that users may require a long time before a natural usage pattern is developed. Our results primarily impact the wearable computing community, but are also relevant in the fields of virtual and augmented reality.

8. REFERENCES