

Gestures to Intuitively Control Large Displays

Wim Fikkert¹, Paul van der Vet¹, Han Rauwerda²,
Timo Breit², and Anton Nijholt¹

^{1,*}Human Media Interaction, University of Twente, The Netherlands
{f.w.fikkert,p.e.vandervet,a.nijholt}@ewi.utwente.nl

²Microarray Department, University of Amsterdam, The Netherlands
{j.rauwerda,t.m.breit}@uva.nl

Abstract. Large displays are highly suited to support discussions in empirical science. Such displays can display project results on a large digital surface to feed the discussion. This paper describes our approach to closely involve multidisciplinary omics scientists in the design of an intuitive display control through hand gestures. This interface is based upon a gesture repertoire. This paper describes how this repertoire is designed based on observations of, and scripted task experiments with, omics scientists.

1 Introduction

A large display is highly suited to support discussions. Such a large digital surface is a valuable resource that can display various pieces of information to feed the discussion. Especially discussions that involve numerous complex visualizations can benefit from such a resource, for example, in empirical science.

Controlling such a large display with a mouse and keyboard is tedious at best. Hand gestures are a powerful means to control the a large display directly. However, display control through hand gestures is not obvious by itself because the large display is an altogether new resource in the discussion. A gesture repertoire is needed that discussants can both understand intuitively and learn easily. We aim to design such a gesture repertoire to control a large display directly through touch and free-handed gesturing. Our gesture repertoire will consist of (using gesture types as defined by Kipp [1]): *deictics* to point to display contents, *iconics* to directly refer to display contents and *beats* to indicate something important. We focus on gestures that are meant to directly control the display; by addressing it explicitly. The goal of this work is to design such a repertoire in close collaboration with its intended users: empirical scientists.

This paper is structured as follows. In Section 2 we introduce omics research as our use case. Section 3 then describes three consecutive experiments in which the gesture repertoire is designed with close involvement of the end-users. Our preliminary results are reported in Section 4. Section 5 discusses the extent to which this repertoire can be applied in other fields.

* We thank our reviewers for their input. This work is part of the BioRange program carried out by the Netherlands Bioinformatics Centre (NBIC), which is supported by a BSIK grant through the Netherlands Genomics Initiative (NGI).

2 Use Case: Omics

A use case in empirical science is highly suited for our study due to complex problems, processes and results. Moreover, diverse expertise is needed to address these problems. Omics is an empirical science in which multidisciplinary research teams address complex biological problems, for example, improving medicines for breast cancer based on genetic expressions in a patient. ‘Omics’ is a suffix that is commonly attached to biological research into the ‘whole’ make-up of an organism on a certain biological level, for example, proteomics on the protein level, and by which huge datasets are produced. An example topic—microarrays—easily generates hundreds of scans that can each contain information about more than 50.000 transcripts. Hence millions of separate measurements have to be analysed. The expertise from all involved disciplines is needed when biological meaning is sought in the experiments’ results. Interviews with these researchers indicated that ‘strange’ results that stand out in the whole can be identified based purely on the result overview. This then leads to a closer analysis.



Fig. 1. The large display in use as a discussion resource

Facilities that support multidisciplinary teams in validating and analysing project results are found in so called dry labs. Such facilities mainly consist of a large amount of computing power and multiple (large) displays. One such dry lab, the *e*-BioLab, is being developed at the University of Amsterdam [2]. One aim of the *e*-BioLab is to enrich omics meetings with more on-demand information to improve their efficiency. Our efforts in the design of a gesture repertoire are aimed at the large display in the *e*-BioLab, see Figure 1. In the *e*-BioLab, we can observe and get involved with scientific discussions that are supported by the large display.

3 Designing the Gesture Repertoire

Large displays have great potential as a discussion resource in face-to-face cooperation. However, such displays are a new phenomenon in discussions; therefore, interacting with them will be a novelty for the participants. An interaction

scheme should be designed with close involvement of our end-users. By studying the behaviour of omics researchers in discussions, an understanding is gained that helps to characterise a gesture repertoire that is tuned to them. We propose three consecutive user studies that gradually build up this repertoire: 1) observing display control in omics research meetings, 2) verifying behaviour cue interpretation through scripted sessions with users and 3) implementing the gesture repertoire in an automated recognition system.

3.1 Experiment 1: Observations in the *e*-BioLab

In the first experiment, we observe discussants using the large display as a discussion resource. The discussions are recorded by four video cameras in addition to the display contents. We then annotate these recordings, focusing on hand gesturing. We aim to identify the minimum information needed for an automated computer vision based system to recognize and interpret these gestures. The following highlights some decisions in our approach.

Recordings in our corpus are reduced to scenes where a discussion between two or more discussants are within arms-length of the display. Gill [3] defines three interaction zones: reflection, negotiation and action. These zones are based on activities in a discussion with a large display resource. Fikkert et al. [4] defined these zones based on the physical distance to the screen: hand-held (action), at arms-length (negotiation) and distal (reflection). The nature of these zones excludes reflection from our annotations, focusing on ‘active’ discussions where discussants react to and interact with the display contents. Discussions in three distinct projects are recorded. We focus on discussions that address the validation and analysis of project results.

Differences in the gestures that are made by users with varying background will mostly be mitigated by the fact that a single chairman heads all of the meetings in the *e*-BioLab. He demonstrates the capabilities of the large display as a discussion resource. We have observed that discussants pick up on these possibilities easily.

Annotations of our recordings consist of transcriptions of person identified speech, body posture, location, gaze direction and hand gestures. We use the Nite XML Toolkit (NXT) [5]. NXT provides annotation tracks in which each modality can be transcribed completely separately and can be synchronised to a common timeline. As such, we are not restricted by a schema that, for example, arrays gestures parallel to speech.

Transcription of hand gesturing requires a notation for written sign language. HamNoSys transcribes body posture, gaze direction and both hand shapes and movements [6]. Languages designed for behaviour synthesis such as MURML [7] and BML [8] do not offer the abstraction capabilities that HamNoSys does. We have used SiGML which is based upon the XML and HamNoSys standards [9]. SiGML can describe scenes on multiple levels: phonology, phonetics and physical articulation. Its XML structure also allows easy incorporation into NXT.

Deictics includes a target on the display. This target is found using both the display’s snapshots and speech annotation [10]. Krandstedt, Kühnlein and

Wachsmuth [7] used a pointing game to study the co-occurrence of speech and pointing. They found that pointing to objects within arms-length was disambiguated through simultaneously occurring speech. We transcribed speech with a focus on specific deictic keywords such as: ‘this’, ‘your’ and ‘it’. These keywords are linked explicitly to gesturing [7,10]. Both body posture and gaze direction are included in the hand gesturing annotations [1]. User location and orientation are found and indicated on a map of the *e*-BioLab using the 4 camera views.

To further classify behaviour, we annotate interaction context and tasks orthogonally based on time occurrences. We distinguish distinct phases in the omics research process: data interpretation, cleaning and quality control. This distinction is based on a previous task analysis in the *e*-BioLab [11].

(Semi-) automated annotation is currently being studied to enrich and speed up the tedious, slow annotation process. Consider computer vision algorithms that (partially) extract body postures or hand positions [12].

Analysis of our annotations aims to identify gestures that the scientists typically make when performing a certain task. Gestures are linked to the orthogonal framework of interaction tasks and context based on their occurrence in time. The difference between the gestures is determined by defining a distance measure, for example, based on the motions and shape of the hands. An unsupervised learning algorithm might be capable of deriving these.

3.2 Experiment 2: Scripted User Sessions

The gestures and their interpretation for controlling the display that were found in the previous experiment are verified here. We ask our end-users to complete a scripted task. These scripted tasks are rooted both in the first experiment and in the task analysis [11]. We examine settings that include one and two users. In the latter case, the users are given roles that suit their experiences.

During this experiment, an operator is in actual control of the system. He has a list of action-reaction cases that was defined in our first experiment. Gestures on this list are described using the annotation scheme from the first experiment. Whenever a participant tries an action that is not on this list, no system response follows which may be confusing to the participant. He then either tries a different action to complete his goal, for example, using a different gesture, or—after some idle time—he is asked explicitly by the operator what his intention is.

These sessions are both recorded and annotated in the same manner as in the first experiment. We aim to verify our basic repertoire by comparing the gesture occurrence and type in similar interaction tasks and context. At this stage, we can add new gestures to, and adjust existing gestures in our repertoire. Interviews with our end-users result give us an appreciation of this new way of large display control.

3.3 Experiment 3: Automated System

This experiment introduces an automated behaviour detection, recognition and tracking system. The operator’s interpretation of a scene may have influenced

the second experiment slightly and it has been removed here. Such an automated system consolidates the gesture repertoire because it cannot profit from human – operator – interpretation of scenes. The aim of this experiment is to polish all previous findings and to arrive at a robust, stable system. Such a system is ideally based on unobtrusive sensing using, for example, computer vision analysis.

A somewhat unobtrusive solution for human behaviour analysis is the best that can be expected in the coming years given the current state of the art in this field [12]. Therefore, we are currently exploring the use of data gloves and a motion capturing system that provide highly accurate behaviour measurements. In the literature it is often argued that obtrusive solutions impose restrictions that influence user behaviour. The extent of this influence, in this setting, remains to be determined.

The result of this last experiment will have confirmed our gesture repertoire. The repertoire will then define the interpretation of gestures in a rule-based manner so that it can be incorporated in an automated system for display control. It can possibly linked to other modalities as well [12].

4 Preliminary Results

Currently, our first experiment is ongoing and we are engaged in our second experiment. Some observations that support and encourage our approach are worth mentioning here. From the start of a meeting, users intuitively move between reflection, negotiation and actions zones as identified by [3]. The reflection zone is used roughly half of the time. Our scientists use the large display actively as a resource in their discussions. When asked, these omics researchers find this resource an ‘indispensable’ asset in their discussions. The fact that other life science groups as well as academic hospitals are building their own *e*-BioLabs supports this opinion. However, our end-users repeatedly indicated that a direct and easy means of controlling the display is needed.

Figure 1 shows end-users using the large display as a discussion resource. Scientists point to pieces of data. They correlate various results that are simultaneously depicted on the display by sequentially ‘grasping’ and walking up to these results whilst arguing their case. Users typically select an object by pointing towards it and then grabbing it. The grabbing gesture varies per user; some use their whole hand, others just linger on the target for a short amount of time. Repositioning an object is done by dragging it to its new location in all cases. Enlarging or shrinking a target does differ to a significant extent; some users use just one hand close to the display by moving their fingers apart, others use their whole arms at arms-length and yet other users grab an object in one hand and resize it by moving their other hand as a virtual sidebar.

5 Discussion

We have described a method to arrive at a gesture repertoire for large display control. Even without the third experiment, we will have constructed a gesture

repertoire that omics researchers can use to operate a large display. Porting our gesture repertoire to other user communities will require further investigation. It seems plausible that, due to the nature of scientific work, this repertoire can be ported to other empirical scientific disciplines. However, generalisation to other user groups may be less easy. Empirical scientists carry out tasks that are highly constrained by both their task environment and their explorative research approach.

References

1. Kipp, M.: *Gesture Generation by Imitation - From Human Behavior to Computer Character Animation*. PhD thesis, Saarland University, Saarbruecken, Germany, Boca Raton, Florida (2004)
2. Rauwerda, H., Roos, M., Hertzberger, B., Breit, T.: The promise of a virtual lab in drug discovery. *Drug Discovery Today* 11, 228–236 (2006)
3. Gill, S., Borchers, J.: Knowledge in co-action: social intelligence in collaborative design activity. *AI and Society* 17(3), 322–339 (2003)
4. Fikkert, W., D'Ambros, M., Bierz, T., Jankun-Kelly, T.: Interacting with visualizations. In: Kerren, A., Ebert, A., Meyer, J. (eds.) *GI-Dagstuhl Research Seminar 2007*. LNCS, vol. 4417, pp. 77–162. Springer, Heidelberg (2007)
5. Carletta, J., Evert, S., Heid, U., Kilgour, J., Robertson, J., Voormann, H.: The NITE XML toolkit: Flexible annotation for multimodal language data. *Behavior Research Methods, Instruments, and Computers* 35(3), 353–363 (2003)
6. Prillwitz, S., Leven, R., Zienert, H., Hanke, T., Henning, J.: *Hamburg Notation System for Sign Languages - An introductory guide*, vol. 5. Signum, Hamburg (1989)
7. Kranstedt, A., Kühnlein, P., Wachsmuth, I.: Deixis in multimodal Human Computer Interaction: An interdisciplinary approach. In: Camurri, A., Volpe, G. (eds.) *GW 2003*. LNCS, vol. 2915, pp. 112–123. Springer, Heidelberg (2004)
8. Vilhjálmsson, H., Cantelmo, N., Cassell, J., Chafai, N., Kipp, M., Kopp, S., Mancini, M., Marsella, S., Marshall, A., Pelachaud, C., Ruttkay, Z., Thórisson, K., van Welbergen, H., van der Werf, R.: *The behavior markup language: Recent developments and challenges* (2007)
9. Elliott, R., Glauert, J., Kennaway, R., Parsons, K.: *Sigml definition*. Technical Report ViSiCAST Deliverable D5-2, University of East Anglia (2001)
10. Kranstedt, A., Lücking, A., Pfeiffer, T., Rieser, H., Wachsmuth, I.: Deixis: How to determine demonstrated objects using a pointing cone. In: Gibet, S., Courty, N., Kamp, J.-F. (eds.) *GW 2005*. LNCS, vol. 3881, pp. 300–311. Springer, Heidelberg (2006)
11. Kulyk, O., Wassink, I.: Getting to know bioinformaticians: Results of an exploratory user study. In: *Workshop on Combining Visualisation and Interaction to Facilitate Scientific Exploration and Discovery in conjunction with British HCI 2006*, pp. 30–38. ACM Press, New York (2006)
12. Pantic, M., Pentland, A., Nijholt, A., Huang, T.: Human computing and machine understanding of human behavior: A survey. In: *ICMI 2006: Proceedings of the 8th international conference on Multimodal interfaces*, vol. 8, pp. 239–248. ACM Press, New York (2006)