A Gesture-Speech Corpus on a Tangible Interface

Dimitra Anastasiou[1], **Kirsten Bergmann**[2]
[1] Luxembourg Institute of Science and Technology (LIST)
5, avenue des Hauts Fourneaux L-4362 Luxembourg
E-mail: Dimitra.Anastasiou@list.lu

[2] Social Cognitive Systems Group, Faculty of Technology (Bielefeld University)
P.O. Box 100 131, D-33501 Bielefeld
E-mail: kirsten.bergmann@uni-bielefeld.de

Abstract

This paper presents a corpus of hand gestures and speech which is created using a tangible user interface (TUI) for collaborative problem solving tasks. We present our initial work within the European Marie Curie project GETUI (GEstures in Tangible User Interfaces). This project involves mainly creating a taxonomy of gestures used in relation to a tangible tabletop which is placed at the Luxembourg Institute of Science and Technology (LIST). A preliminary user study showed that gesturing encourages the use of rapid epistemic actions by lowering cognitive load. Ongoing corpus collection studies provide insights about the impact of gestures on learning, collaboration and cognition, while also identify cultural differences of gestures.

Keywords: human-computer interaction, pointing, tangibles, taxonomy

1. Introduction

Gesturing is a natural communication means with both inter-personal and intra-personal functions. Inter-personally, in human face-to-face interaction (HHI), co-speech gestures emphasize or supplement spoken content. Intra-personally, gestures can support cognitive processing (e.g. Ping & Goldin-Meadow, 2010), a fact which can be exploited for so-called Tangible User Interfaces (TUIs). The term TUI has been established by Ullmer & Ishii (2000) as follows: "[TUIs] give physical form to digital information, employing physical artifacts both as 'representations' and 'controls' for computational media. TUIs "[provide] tangible representations to digital information and controls, allowing users to quite literally grasp data with their hands" (Shaer & Hornecker, 2010). Kirk et al. (2009) stated that the kinesthetic memory of moving a tangible object can increase the recall of performed actions, preventing mode errors, as interacting with a physical object can be equivalent to an implicit, user-maintained mode. Likewise, Esteves et al. (2013) showed that conducting problem solving tasks on a TUI encourages rapid *epistemic* actions, and lowers cognitive load by simplifying thinking processes. Accordingly, TUIs are of particular interest for the domain of technology-enriched learning environments, especially given the option to integrate technology-based assessment (TBA). However, currently there is neither a systematic analysis of employing TUIs in the context of TBA, nor has interaction with TUIs been systematically explored in TBA yet.

In this paper we present first steps towards a detailed investigation of integrating TUIs and TBA. Our goal is to set up a data-based taxonomy of gestures used in interaction with a TUI, whereby our domain of application is collaborative problem solving as one of the most important 21st Century skills. The paper is laid out as follows: Section 2 presents related work on gesture taxonomies in general as well as the very relation of gesture and cognition as well as gesture and culture. In Section 3 we present a pre-study and its results. The design of our corpus collection studies is described in Section 4, followed by draft for the gesture taxonomy. Finally, we discuss annotation issues and draw conclusions with respect to the impact of the corpus.

2. Related Work

Gesture taxonomies have been presented in the literature both from a philological viewpoint and a human-computer interaction (HCI) viewpoint. Regarding the former, foundations about gestures were established by McNeill (1992), based on Kendon's continuum (Kendon, 1982); gestures were classified into *gesticulation*, *pantomime*, *emblem*, and *sign language*. *Gesticulation* is further classified into *iconic*, *metaphoric*, *rhythmic*, *cohesive*, and *deictic* gestures. As for the latter, from an HCI perspective, Quek (1994) created a taxonomy in HCI, classifying meaningful gestures into *communicative* and *manipulative* gestures. Manipulative gestures can occur either on the desktop in a 2-D interaction using a direct manipulation device, as a 3-D interaction involving empty-handed movements to mimic manipulations of physical objects, or by manipulating actual physical objects that map onto a virtual object in TUIs. We focus particularly on the third categorization of manipulative gestures. The most prevalent type of gesture in relation to TUIs is *pointing* or *deictic*. Moreover, Lao et al. (2009) defined *tapping*, *pressing*, and *dragging* gestures and showed that a variety of hand gestures can be constructed through these three basic movements. Karam and Schraefel (2005) made a classification of the literature about gesture interaction research (mainly user studies) since the early 1990s.

As far as hand gesture recognition is concerned, Rautaray & Anupam (2015) made a recent survey on vision-based hand gesture recognition for HCI, analysing the three main recognition techniques (detection, tracking, recognition) as well as the required software platforms.

Gesture and cognition Alibali et al. (2000) stated that gesturing reduces the cognitive load for both adults and children, particularly during explanation tasks. Klemmer et al. (2006) pointed out that systems that constrain gestural abilities, e.g. having the hands stuck on a keyboard, are likely to hinder users' thinking and communication.

Gesture and culture Gestures and their cultural connotations have been examined, among others, by Archer (1997) and Kita (2009). Archer (1997) found that there are both cultural differences and meta-differences, i.e. more profound differences involving deeply embedded categories of meaning that make cultures unique. Kita (2009) reviewed the literature on cross-cultural variation of gesture based on four relevant factors: conventions of form-meaning association, language, spatial cognition, and pragmatics of gesture use. Moreover, we had previously defined a *locale* as a combination of language and culture as well as gesture localization as follows: *Gesture localization is the adaptation of gestures to a target locale in order to transfer the same meaning as in the original locale* (Anastasiou, 2011). The importance of addressing cultural differences in gesture use becomes more and more important in times of globalization and migration. In Luxembourg, for instance, there were 220,522 foreigners equivalent to 43.04% of the total population in 2011[1]. In this multi-lingual and -cultural society, it is essential to be aware of gestures based on other cultures, so that humans and machines communicate 'properly' by respecting other people's culture.

In our research, we aim to integrate all the aforementioned aspects by setting up and analyzing a corpus of speech-gesture use in collaborative interaction with a TUI. The goal is to create a gesture taxonomy both from a philological and HCI perspective under consideration of cognitive skills and cultural differences of gesture use.

3. Pre-Study

A preliminary user study was conducted at the Luxembourg Institute of Science and Technology (LIST) (Anastasiou et al., 2014). There were 10 groups of three people in each group; the task of the participants was to explore the relation of external parameters on the production of electricity of a windmill presented on a tangible tabletop. The goal of the study was to observe, analyze and understand the interactions between participants while collaboratively solving a task. We annotated in total 601 gestures, 334 of which were manipulative, 181 pointing, followed by 35 emblems, 28 iconic, and 23 adaptors. The gesture analysis based on this preliminary study resulted in the following gesture taxonomy:

1. ***Deictic/pointing*** gestures: point something/somewhere, such as to a(n):
    a. Object(s);
    b. TUI;
    c. First object and then TUI;
    d. Other participant(s);
    e. Collaborative pointing.
2. ***Iconic*** gestures: indicate distance, depth, or height or describe the shape of an object;
    a. *Encircling*: making a circle with fingers representing the turning of the physical object;
    b. *Moving an open hand forward/backward*: representing distance and/or asking from a participant to move the physical object forward/backward;
    c. *Moving an open hand downwards vertically*: representing depth.
3. ***Emblems***: have a direct verbal translation and can be interpreted differently by different cultures;
    a. *Holding open hand*: prompting other participants to wait or stop interaction;
    b. *Raising hand with palm up*: indicating uncertainty, questioning "what are we/you doing?";
    c. *Showing an open hand:* prompting other participants to continue interaction;
    d. *Raising finger/arm (open hand)*: indicating uncertainty, such as "I do not know";
    e. *Shaking fingers in a circular way*: indicating fuzziness, like "so and so".
4. ***Adaptors****:* are not used intentionally during a communication or interaction;
    a. Head/chin/nose scratching;
    b. Touching nose/mouth.
5. ***TUI-related/manipulative*** gestures: occur specifically in interaction with TUIs.
    a. *Placing*: taking the object from table frame and putting it on a specific position on the TUI;
    b. *Tracing*: moving the object to another place of the table by dragging it on the TUI;
    c. *Rotating*: turning the object from right to left or left to right;
    d. *Moving*: holding up the object from table and placing it somewhere else on the TU.

In general, our study showed that problem solving task on the TUI encouraged the use of rapid epistemic actions by simplifying thinking processes. This conclusion is drawn by two results of our study: (1) almost half of the gestures were not TUI-related, i.e. did not modify anything in the simulation and just helped to lower cognitive load by simplifying the thinking process (Esteves, 2013) and (2) in case of a gesture, the other participants reacted also with gestures (85.4% TUI-related gestures); this shows that modifications on the parameters could be quickly done and feedback was provided immediately. Moreover, we observed that 78,5% spoke during gesturing, which shows the tight connection between speech-gesture, as already well established in the literature (Goodwin, 1994; Ping & Goldin-Meadow, 2010).

4. Corpus collection studies

Participants To take cultural differences into account, we recruit 60 participants in our evaluation studies, separated in 3 *locales*: 20 francophone, 20 germanophone, and 20

[1] Statistics Portal: http://www.statistiques.public.lu/en/news/population/population/2012/08/20120821/index.html, 18.02.16

anglophone. The participants are minor students (15-18 years old) and recruited through public schools in Luxembourg.

Task Participants' task is based on a microworld; the three pupils will be provided with three physical objects that represent industrial facilities that produce electricity, e.g. a windmill, photovoltaics and a coal-fired power station. The objects will be given artificial names or variables in order to avoid previous knowledge. By turning the objects on the TUI (input: 0-10 scale), there are two parameters changing: i) the electricity generation and ii) CO2 emission. The pictures depicted on the tabletop will be accordingly adapted to the output values. This task is similar to tasks given in the international large-scale educational Programme for International Student Assessment (PISA) programme.

Study Setup The TUI employed for the study at LIST institute in Luxembourg is realised as a tangible tabletop (75x120 cm). Physical objects can be manipulated on the table in order to explore different factors. The table provides visual feedback in real-time and displays the effects with pictures and animations.

Data A multimodal corpus of video volume≈9 h and 200GB is currently collected. The Kinect 2.0[2] depth sense camera is used for recognition of the spatial position of the participants, proximity (between users and between users and tabletop) and their gestures. The light-weight and extensible software framework TULIP (Tobias et al. 2015) is used, which combines the TUI interaction paradigm and software engineering principles. We will draw upon the collaborative problem solving assessment approach that was employed in PISA 2015. We will follow the MicroDYN framework of Greiff et al. (2012), a new approach for computer-based assessment of CPS based on linear structural equations. This methodology allows to formally describe everyday activities by means of variables, outcomes and their interconnectedness.

4.1 Annotation challenges

Our user studies will result in a multimodal corpus of speech and gesture that will be annotated with ELAN (Wittenburg et al. 2006) and NEUROGES (Lausberg & Sloetjes, 2015). It will be examined whether speakers of typologically different languages exhibit differences in their gestural patterns, how gestures are coordinated with intonation and to which degree are semantically and pragmatically co-expressive with the verbal utterance. Moreover, we will examine which part-of-speech (PoS) users used in every gesture phase of the gesture unit: *preparation, stroke,* and *retraction.* (Kipp, 2004). For instance, for the sentence "This belongs here", they might use the pointing gesture synchronously with the word *this* or the word *here* or they might use two subsequent gestures. Our gesture taxonomy will include *locale*-specific gestures, as participants from three different locales are recruited.

In addition, in our evaluation studies it is examined whether the spatial context/proximity affects the gestural performance. For example, we observe whether it plays a role where exactly the participant stands in relation to the physical objects or the objects in relation to the tabletop. Through these dimensions, we assess collaborative complex problem solving and reasoning skills.

4.2 Gesture taxonomy

Our gesture taxonomy is partially based on our draft taxonomy (see Sec. 3) with the difference of putting together *pointing* and *iconic* gestures under *physical,* and *emblems* and *adaptors* under *affective* gestures. Moreover, this taxonomy is extended by adding the category of *collaborative* gestures. Because of page constraints, only a few layers are presented here.

1. *Physical*
    a. *Pointing*
        i. *Single-handed*
            - *Object*
            - *TUI, etc.*
        ii. *Bimanual*
    b. *Iconic*
2. *Manipulative*
    a. *Placing*
    b. *Removing*
    c. *Rotating*
    d. *Tracing*
3. *Affective*
    a. *Emblems*
    b. *Adaptors*
4. *Collaborative*
    a. *Symmetry*
        i. *Symmetric*
        ii. *Asymmetric*
        iii. *Partially symmetric*
    b. *Parallelism*
    c. *Additivity*

The last category is cross sectoral, as it combines gestures from other categories. Collaborative gestures have been examined in the literature by Block et al. (2015), Tang et al. (2006), and Morris (2006). Morris (2006) stated that symmetry axis refers to whether participants in a cooperative gesture perform identical actions or distinct actions, while parallelism is defined as the relative timing of each contributor's axis. An *additive* gesture is one which is meaningful when performed by a single user, but whose meaning is amplified when simultaneously performed by all members of the group.

Very often in the literature about gesture taxonomies in HCI, gestures are mixed with verbal utterances. In our annotation scheme, speech is considered as a separate modality and will be first transcribed, then annotated based on Conversational Analysis and third examined in temporal coordination with gestures. Verbal utterances are categorized into substantial or pragmatic (Kendon, 2004), interpretation, conflicts, negotiation, etc. We plan to extend the CPS model of dialogue by Blyloke and Allen (2005) and Hmelo-Silver (2003).

5. Conclusion

The research presented here addresses a practical application field of HCI and Interaction Design: TUIs. In the literature it has been shown that gesturing can lower cognitive load, a fact that we also substantiated in our pre-study. The main objective of our current research is to

[2] http://www.xbox.com/en-US/xbox-one/accessories/kinect-for-xbox-one

explore through user studies the gestural performance of users while interacting on a TUI in a collaborative problem solving task and in addition, what kind of effect does this the gestural performance have on 21th Century skills. We explore these aspects through a corpus collection study with 60 pupils from three different locales. The data will be analysed with respect to how participants interact with each other (gestures from HHI perspective) and with objects/TUI (HCI perspective) and how these both kinds of interaction facilitate the technology-based assessement.

As far as the impact of corpus is concerned, both gesture and speech researchers will benefit from its existence and annotation. Moreover, educators and pupils will learn not only about power grids, but at a more abstract level, about computer-mediated collaborative problem solving in general. At a higher level, we will provide guidelines for future applicability of TUIs in PISA. Moreover, as we will have 3D data as output of the Kinect, gesture recognition researchers can train their systems and increase the accuracy, particularly for hand and finger gestures which is a quite big recognition challenge. Particularly, such a scenario is particularly challenging, as there are many users crossing in front of each others or placing hands on the top of other hand(s).

6. Acknowledgements

This research is funded by the Marie Curie Individual Fellowship project *GEstures in Tangible User Interfaces*.

7. Bibliographical References

Alibali, M.W., Kita S, Young A. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language & Cognitive Processes*, 15, pp. 593--613.

Anastasiou, D., Maquil, V., Ras, E. (2014). Gesture Analysis in a Case Study with a Tangible User Interface for Collaborative Problem Solving. *Journal on Multimodal User Interfaces*, Springer.

Anastasiou, D., (2011). Speech Recognition, Machine Translation and Gesture Localisation. *TRALOGY: Translation Careers and Technologies: Convergence Points for the Future,* 3 - 4 March, Paris, France.

Archer, D. (1997). Unspoken diversity: cultural differences in gestures. *"Visual Sociology" Qualitative Sociology*, 20(1), pp. 79--105.

Block, F. et al. (2015). Fluid Grouping: Quantifying group engagement around interactive tabletop exhibits in the wild. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015, pp. 867--876.

Blyloke, N., Allen, J., (2005). A collaborative problem-solving model of dialogue. *Proceedings of the SIGdial Workshop on Discourse and Dialog*.

Esteves, A. et al. (2013). Physical games or digital games?: Comparing support for mental projection in tangible and virtual representations of a problem solving task. *Proceedings of TEI*, pp. 167--174.

Goodwin, C. (1994). *Professional vision*. American Anthropologist, 96(3), pp. 606--633.

Greiff, S., Wüstenberg, S., Funke, J. (2012). Dynamic Problem Solving: A new measurement perspective. *Applied Psychological Measurement*, 36, pp. 189--213.

Hmelo-Silver, C.E. (2003). Analyzing collaborative knowledge construction: multiple methods for integrated understanding. *Computers & Education* 41, pp. 397--420.

Karam, M., Schraefel, mc(2005). *A Taxonomy of Gestures in Human Computer Interactions,* Technical Report.

Kendon, A. (1982). The study of gesture: some observations on its history. *Rech Semiot Semiot Inq* 2(1), pp. 25--62.

Kendon, A. (2004). *Gesture-Visible Action As Utterance*, UK: Cambridge University Press.

Kirk, DS. et al. (2009). Putting the physical into the digital: issues in designing hybrid interactive surfaces. *Proceedings of BCS HCI 2009,* pp 35--54.

Kipp, M. (2004). *Gesture generation by imitation—from human behavior to computer character animatio*n. PhD Dissertation, Boca Raton, Florida.

Kita, S. (2009) Cross-cultural variation of speech-accompanying gesture: A review. *Language and Cognitive Processes*, 24(2), pp. 145--167.

Klemmer, S.R., Hartmann, B., Takayama, L. (2006). How bodies matter: Five themes for interaction design. *Proceedings of DIS 2006 Conference on Designing Interactive Systems*, pp. 140--149.

Lao, S. et al. (2009). A gestural interaction design model for multi-touch displays. *Proceedings of the British HCI-Group*.

Lausberg, H., Sloetjes, H., (2015). The revised NEUROGES-ELAN system: An objective and reliable interdisciplinary analysis tool for nonverbal behavior and gesture. *Behavior Research Methods*.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* Chicago: University of Chicago Press.

Morris, M.R. et al. (2006). Cooperative gestures: multi-user gestural interactions for co-located groupware. *Proceedings of CHI 2006*.

Ping, R., Goldin-Meadow, S. (2010). Gesturing saves cognitive resources when talking about nonpresent objects. *Cognitive Science,* 34, pp. 602--619.

Quek, F. (1994). Toward a vision-based hand gesture interface. *Proceedings of the virtual reality, software and technology conference*, pp. 17--31.

Rautaray, S., Anupam A. (2015). Vision based hand gesture recognition for human computer interaction: a survey. *Artificial Intelligence Review,* 43(1), pp. 1--54.

Shaer, O., Hornecker, E. (2010). Tangible user interfaces: past, present and future directions. *Found Trends Hum Comput Interact,* 3 (1–2), pp.1--137.

Tang, A. et al. (2006). Collaborative coupling over Tabletop Displays. *Proceedings of CHI 2006.*

Ullmer, B, Ishii, H. (2000) Emerging frameworks for tangible user interfaces. *IBM Systems Journal*, 39, pp. 915--931.

Wittenburg, P. et al. (2006). ELAN: a professional framework for multimodality research. *Proceedings of the 5th Conference on Language Resources and Evaluation*, pp. 1556--1559.