Unraveling Students' Interaction Around a Tangible Interface Using Gesture Recognition

Bertrand Schneider
Stanford University
schneibe@stanford.edu

Paulo Blikstein
Stanford University
paulob@stanford.edu

ABSTRACT

In this paper, we describe techniques to use multimodal learning analytics to analyze data collected around an interactive tangible learning environment. In a previous study [4], we designed and evaluated a Tangible User Interface (TUI) where dyads of students were asked to learn about the human hearing system by *reconstructing* it. In the current study, we present the analysis of the data collected in form of their gestures, and we describe how we extracted meaningful predictors for students' learning from this datasets. We explored how Kinect™ data can inform "in-situ" interactions around a tabletop (i.e. using clustering algorithms to find prototypical body positions). We discuss the implications of those results for analyzing data from rich, multimodal learning environments.

Keywords

Tangible User Interface; Data Mining; Constructionism.

1. INTRODUCTION

Students' gestures have been extensively researched in the learning sciences: numerous studies on embodied cognition have unraveled links between learning and students' gestures. More generally, there has been a plethora of studies about people's intuitive representations of everyday situation and bodily language [3]. This line of research has provided new ways to understand the way students integrate new concepts to their everyday understanding of science phenomena.

Yet, this field of research suffers from serious methodological limitations. Most studies are qualitative by nature, or require researchers to manually annotate hours of video recordings. Now that the theoretical underpinnings of the field are established, it would be the right time to speed up discovery and data analysis, but the pace at which new results are generated is slower than desired, especially for highly granular data. The emerging field learning analytics and educational data mining, and especially the field of multimodal learning analytics [9] might provide just the right data collection and analysis tools to tackle this problem. Thus, the goal of this paper is to address this methodological gap by suggesting new ways to conduct research on students' body language, as well as providing news lenses to look at students' micro-behaviors while learning. In our study, we collected data on user's actions by using a Microsoft Kinect™. We then used data mining techniques to make sense of those two datasets.

2. THE CURRENT STUDY

In a previous study, we were interested in pursuing the work started with two other TUIs developed in our lab. In our research, we have found that TUIs can be advantageously used in a discovery-learning situation when students approach an unfamiliar topic compared to a standard "tell-and practice" instruction (i.e. using a TUI before, rather than after, a standard kind of instruction such as reading a textbook chapter or attending a lecture). In a controlled experiment [6], we showed that students who first used a TUI and then read a textbook chapter outperformed students who completed the same activities but in the reverse order (text followed by TUI). To show that being physically engaged doesn't fully explain our results, we designed the following experiment, where students were asked to discover how the human hearing system works (N=38). Pairs of students worked on a tangible interface called EarExplorer (Fig. 1) where they learned about the human hearing system by reconstructing it. In one condition, students rebuilt the hearing system by trial and error, using resources provided by the system. In a second condition, they used the same setup except that a video of teacher demonstrated how to rebuild the hearing system and explained the function of each organ as students progressed through the activity. Students in both conditions then read a textbook chapter explaining sound transduction in a more formal way. We found that students in the first group achieved a higher learning gain as measured by the pre and post-test. A MANOVA showed that participants in the "discover" group learned significantly more after the first activity: $F(1,35) = 22.11$, $p < 0.001$ and after the second activity $F(1,35) = 16.15$, $p < 0.001$ compared to the participants in the "listen" condition.

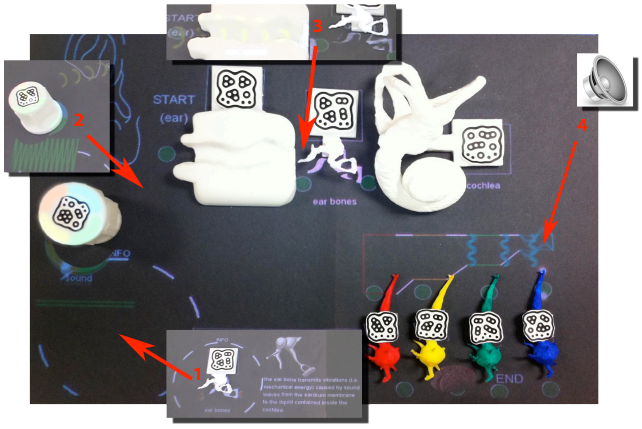

Figure 1: The EarExplorer Interface. Students use the infobox (1) to learn about the different organs; they then generate sounds at different frequencies with a speaker (2); sound waves travel from the emitter through the ear canal to the ear bones (3); finally, the sound reached the basilar membrane inside the cochlea, activates a specific neuron and replayed the sound if the configuration is correct (4).

Those results suggest that hands-on activities alone don't fully take advantage of educational TUIs; rather, discovery should be an integral part of designing interactive hands-on activities. Given those findings, our goal is to take a new look at this dataset by applying data mining and multimodal learning analytics techniques to students' body language: we will analyze the logs generated by a Kinect™ sensor that recorded students' actions.

2.1 Kinect data

As described earlier, a Kinect sensor captured the body movements of both students during the learning activity (30 data points per second). For the sake of simplicity, we only analyzed the students' body language during the hands-on activity (i.e. when reconstructing the hearing system, as opposed to reading the text which is the second activity). This step lasted 15 minutes, which gives us 15 * 60 * 30 = 27000 data points for each participant. Since 38 students took part in the study and two were removed for missing data, we have approximately 1 million data points from the Kinect sensor. This is a relatively large dataset that needs to be drastically simplified in order to make sense of it.

2.2 Movements

The most straightforward (and admittedly naïve) approach to analyzing the Kinect dataset is to compute the amount of movement generated by each participant. The hypothesis is that more engaged students move their body more than less engaged ones, and that physical movement is a proxy for general engagement; this general engagement in turn is related to learning gains. There are two ways to compute this metric: the first one is by calculating the Euclidean distance between each joint and averaging the result over time. This approach is not ideal, because it does not take into account the natural variations in students limbs' lengths. An arguably better way to compute movements is to look at variations in angles between joints in body positions. We tried both approaches and sliced the data over time to get a measure for each minute. We also computed an overall score, as well as a score for each joint. We did not find any significant correlation between the measures described in this paragraph and learning gains. For instance the amount of movement computed with joint angles produced the following correlation: $r(34) = 0.079$, $p = 0.648$. On the one hand, this result is somewhat surprising: we would expect at least some of those measures to be associated with higher engagement and thus more learning. On the other hand, a movement of the hand, for instance, can mean a range of different things (e.g. a sign of boredom, interest, a deictic gesture, and so on), so ultimately the results make sense. Many simple gestures are ambiguous by nature, and in our particular case we did not have enough information to correctly contextualize them. In the next sections, we look at more refined measures of students' activity, such as bimanual coordination, body synchronization and postures.

2.3 Bimanual Coordination

In a related study, Worsley & Blikstein [9] have shown that bimanual coordination was predictive of participants' expertise in solving an engineering problem. Based on these results, we decided to compute a similar metric. More specifically, the idea is to compute and compare the amount of movement generated by each hand. Figure 2 shows two examples generated by this approach: on the right, the student barely used his left arm and the student on the left is used both arms during the entire activity. However, to make sense of this metric, we need to introduce additional results that we found on the initial dataset.

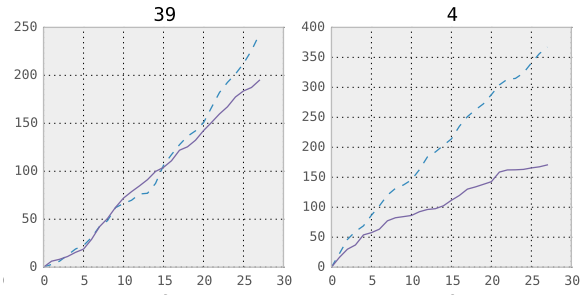

Figure 2: hand coordination of two students during the activity. Participant #39 is bimanual and uses both hands. Participant #4 uses predominantly his right hand (blue line).

Previous research has shown that each student working in groups can often be categorized as either being the "driver" or the "passenger" of the interaction [7]. Several indicators can be used to categorize each dyad's members: 1) who started the discussion when the experimenter leaves, 2) who spoke most, 3) who managed turn-taking (e.g., by asking "what do you think?", "how do you understand this part of the diagram?"), and 4) who decides the next focus of attention (e.g., "so to summarize, our answers are [...]. I think we need to spend more time on X"). This measure can be considered as an aggregate estimation over the whole activity of the dyad's dynamic profile. We acknowledge that subjects are likely to shift roles during the activity. We also recognize that this categorization is more likely to be a continuum, and that in a few cases the difference between drivers and passengers may be subtle.

In our case, after making this distinction for students, we further separated them by computing a median-split on their GPA. This resulted in four categories: a student could either be a driver or a passenger with a high or low GPA (Fig. 3). Surprisingly, having a proficient driver in the group does not lead to higher learning gains: $F(1,16) = 0.04$, $p = 0.84$, Cohen's $d = 0.17$ (low GPA driver: mean=7.63, SD=1.84; high GPA driver: mean=7.87, SD=2.57). On the other hand, having a passenger with a high GPA *does* lead to increased learning gains: $F(1,18) = 3.51$, $p = 0.08$, Cohen's $d = 1.4$ (low GPA passenger: mean=6.22, SD=2.26; high GPA passenger: mean=8.36, SD=2.43).

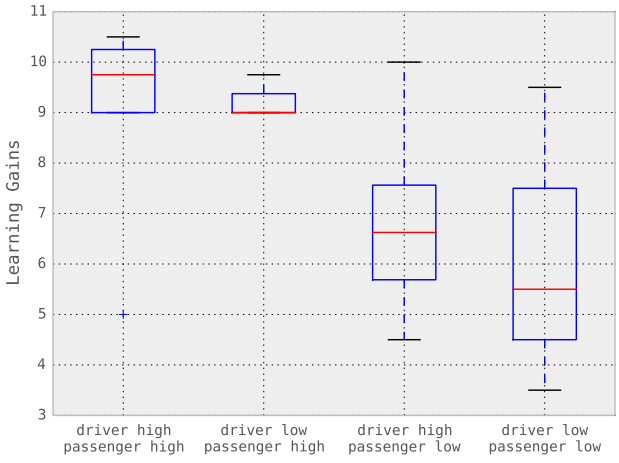

Figure 3: Boxplots of the four kinds of dyads described above: driver / passenger with high / low GPA.

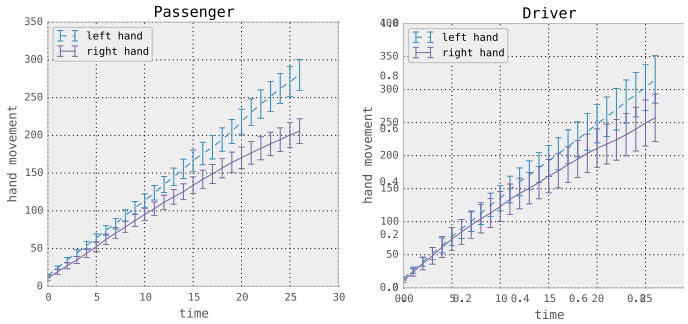

Figure 4: Bimanual coordination from Drivers and Passengers in dyads of students.

This result is not totally unexpected: proficient students who do not "take control" of the activity tended to leave more space for trial and error to their partner and suggested hints when needed. This situation resulted in increased participation and engagement from the low GPA student. In the opposite situation, the same student would stay passive and let the driver solve the problem on her/his own. The distinction between proficient / non-proficient drivers / passengers allowed us to find interesting patterns in our data. More specifically, we found that drivers tend to use both hands while the passenger uses at most one hand. Figure 4 show the aggregated evaluation over time of the hand movements of those two types of students. Using an ANOVA, we found a significant differences between the amount of movements of each hand for the passengers at the end of the activity: $F(1,35) = 7.66$, $p = 0.01$ (left hand mean=280.00, SD=86.30; right hand mean=205.55, SD=69.62). This difference was not significant for the drivers: $F(1,35) = 1.24$, $p = 0.27$ (left hand mean=315.38, SD=152.32; right hand mean=257.21, SD=152.27).

This result shows that we can potentially discriminate between drivers and passengers by looking at their hand movements. As a possible implication of this result, we can imagine future systems where machine-learning algorithms will make predictions about the "status" of each member of a dyad. Using many more features, we can imagine a learning environment where personalized scaffolding is provided depending on the groups' dynamic: proficient leaders can be encouraged to take a more passive role, while less proficient students would be provided with more scaffoldings and more opportunities to participate.

2.4 Going Beyond Movements

While the previous results provide us with interesting metrics to predict learning, they are rather unsophisticated. In this section we describe how we used clustering algorithms to automate the creation of coding schemes on body postures. Recall that most previous work in this area was conducted manually, by analyzing videos frame by frame in a highly time consuming process. If we can show that an algorithm can accomplish a similar task, it will provide researchers with an easy and efficient way to quickly analyze students' body language. Our approach was to take our entire dataset (1 million entries), and transform it into (joint) angles instead of positions in a three-dimensional Cartesian coordinate system. We then fed this matrix into a simple clustering algorithm (K-means) that provided us with prototypical body positions. As a first step, we decided to keep our analyses as simple as possible and limited the number of clusters to three. The results are shown in Figure 5. We found the three clusters to have interesting properties: the first one (top left) represents an "active" position: both arms are on the table, supposedly manipulating something or at least ready to act; the head is tilted toward the table in an attentive position. The second cluster (top right) shows a "semi-active" posture: one arm is flexed, while the other one is straight on the table, probably manipulating a tangible. The last one (bottom left) represents a "passive" posture, where both arms are crossed and the body looks relaxed. We then used those three clusters to classify each data point into one of those three clusters based on proximity to cluster centroids and counted how many times each student was in each posture. The way we interpret those three clusters seem to correlate with our previous measures: the first posture is positively associated with students' learning gains $r(34) = 0.329$, $p < 0.05$ while the third one is negatively correlated with students learning gains $r(34) = -0.420$, $p < 0.05$. Additionally, we found that the number of times students transitioned from one posture to another was also significantly correlated with their learning gains: $r(34) = 0.335$, $p < 0.05$. This suggests not only that some postures are indicative of learning, but certain sequences of postures are too.

Previous work [8] has shown that "ideal" cycles of cognition (i.e. planning, executing and evaluating an action) are usually associated with higher performances and higher learning gains. It is possible that the results of our clustering algorithm produced a similar construct: an increased number of cycles where students think for a while (posture 1 and 2) and then execute an action (posture 3) could be interpreted as something akin to an ideal cycle of cognition described by [8]. We should mention that we tried several approaches before finding the optimal way to cluster our dataset. We first tried to use joint *positions* in a three-dimensional space, as measured by the Kinect (i.e. the x,y,z coordinates of each joint of the kinect skeleton: head, neck, shoulders, elbows, arms). We found two main issues with this method: first, clusters were influenced by students' orientation toward the tangible interface (right or left side). Second, the size of their limbs interfered with the clustering algorithms: longer limbs were more likely to be clustered together.

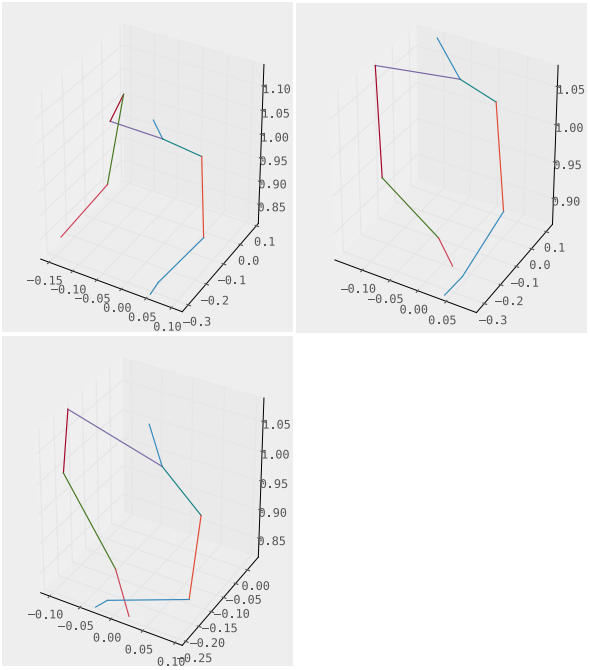

Figure 5: The results of the k-means algorithm on students' body posture (1 million data points).

2.5 Analyses at the dyad level

We describe here additional results conducted at the dyad level, i.e. when taking both bodies into consideration.

2.5.1 Body Synchronization

In a previous study [5], Schneider & Pea found that students' visual synchronization (as measured by eye-trackers) was correlated with their learning gains. That is, more moments of joint attention was beneficial to establishing a common ground, which in turn positively influenced how much students learned during an activity. Other lines of research suggest that body synchronization is associated with productive collaborations [1]. We were inspired by those results and decided to compute a metric for gestures synchronization using the Kinect data. Our approach was to first take pairs of data points (one from each student) and computes the distance between them. Distance was calculated by taking the absolute value of the difference between the joint angles of each participant. An ANOVA did not reveal any significant effect of this measure on our experimental manipulation: $F(1,17) = 0.92$, $p = 0.35$, Cohen's $d = 0.14$ ("discover" mean=0.46, SD=0.09; "listen" mean=0.42, SD=0.05). We also did not find a significant correlation between body synchronization and learning gains: $r(16) = 0.189$, $p = 0.453$. It suggests that even though gaze synchronization is a strong predictor for students' quality of collaboration, body synchronization does not hold the same properties in our dataset.

2.5.2 Body Distance

A last metric that we assessed was inspired by the theory of Proxemics developed by Edward T. Hall [2]. In this seminal work, he proposed to categorize the distances around a person into different zones: the intimate area (less than 15 cm to 46 cm), the personal space (46 to 122 cm), the social distance (122 to 370 cm) and the public distance (370 to 760cm or more). Interestingly, in our study students were seated at a distance that varied between the intimate and personal distance. Moving from a personal to an intimate distance is considered a violation of someone's territory if there is not implicit agreement that someone can do so. Thus, a small distance between two students can potentially characterize a productive collaboration and thus higher learning gains. Similarly, a larger distance can be an indicator of a poor collaboration. We computed the distance between students at each data point (i.e. 30 times per second) by taking the rightmost joint from the student on the left side and the leftmost joint from the student on the right side of the table; we then calculated the Euclidean distance between those two points and averaged a global score for the entire activity (27000 data points). We did not find a correlation between learning and the distance between students' bodies: $r(16) = 0.377$, $p = 0.123$. However we found that this metric was correlated with students' pre-existing knowledge on the topic taught (i.e. score on the pre-test): $r(16) = -0.548$, $p = 0.019$. While there could be multiple interpretations of this result, it suggests that students who are unfamiliar or uncomfortable with the subject matter tend to establish a larger distance with their peers and possibly be more defensive during a collaborative activity.

3. DISCUSSION

In this paper, we developed several metrics that can be used to predict students' learning around an interactive tabletop. We found that the raw amount of movement was not a relevant predictor for our purposes; however, bimanual coordination was predictive of students' leadership in a group. Additionally, clustering body position with k-means provided us with three categories: we found that "active" positions were correlated with learning gains, "passive" positions were negatively correlated with learning gains, and that transition between those states was predictive of learning. Third, we explored students' body language on a social level: contrary to common social psychology theories, we found that body synchronization was not correlated with any of our measures. Fourth, the distance between students' bodies during the activity was associated with their pre-existing knowledge on the topic taught: students with low scores on the pre-test tended to be further away from their partner compared to students who obtained a high score.

4. CONCLUSION

Our goal was to show the potential of rich datasets for advancing our understanding of students' learning trajectories. We contrast our approach with online settings where the data is drastically more limited (e.g. click-stream data). We believe that insights are more likely to be generated in these "in-situ" settings, because researchers can more easily collect relevant educational data: for instance gestures captured with a Kinect sensor, eye gaze movements collected with (mobile) eye-trackers, and arousal measures gathered using galvanic skin response sensors. There are multiple implications stemming from our results. One of them is that seemingly erratic events such as gestures can be objectively correlated with learning gains in ecologically-valid tasks -- i.e., students were using an interactive tabletop which had no constrains for gestures -- everything was allowed. The task itself was very open-ended -- there were multiple paths to success. This is a departure from research that uses data mining in very constricted and well-structured tasks.

5. REFERENCES

[1] Chartrand, T.L. and Bargh, J.A. The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology 76*, 6 (1999), 893–910.

[2] Hall, E.T. *The hidden dimension (1st ed.)*. Doubleday & Co, New York, NY, US, 1966.

[3] Roth, W.-M. Gestures: Their Role in Teaching and Learning. *Review of Educational Research 71*, 3 (2001), 365–392.

[4] Schneider, B., Bumbacher, E., Salehi, S., & Blikstein, P. Discovery Learning versus Traditional Instruction with a Tangible Interface. *Unpublished Manuscript.*

[5] Schneider, B. and Pea, R. Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *International Journal of Computer-Supported Collaborative Learning 8*, 4 (2013), 375–397.

[6] Schneider, B., Wallace, J., Blikstein, P., and Pea, R. Preparing for Future Learning with a Tangible User Interface: The Case of Neuroscience. *IEEE Transactions on Learning Technologies 99*, 1 (5555), 1.

[7] Shaer, O., Strait, M., Valdes, C., Feng, T., Lintz, M., and Wang, H. Enhancing genomic learning through tabletop interaction. *In Proc. of the 2011 conference on Human factors in computing systems*, ACM (2011), 2817–2826.

[8] Tschan, F. Ideal Cycles of Communication (or Cognitions) in Triads, Dyads, and Individuals. *Small Group Research 33*, 6 (2002), 615 –643.

[9] Worsley, M. and Blikstein, P. Towards the Development of Multimodal Action Based Assessment. *In Proc. of LAK2013*, ACM (2013), 94–101.