Blink: Observing Thin Slices of Behavior to Determine Users' Expectation Towards Task Difficulty

Nuno Branco
School of Technology and Management of Felgueiras / engageLab, University of Minho
Felgueiras, Portugal
nuno@engagelab.org

João Pedro Ferreira
engageLab, University of Minho
Guimarães, Portugal
jpferreira@engagelab.org

Marta Noronha e Sousa
engageLab, University of Minho
Guimarães, Portugal
msousa@engagelab.org

Pedro Branco
engageLab / Dep. Inf. Sys., University of Minho
Guimarães, Portugal
pbranco@dsi.uminho.pt

Nuno Otero
Algoritmi Centre, University of Minho
Guimarães, Portugal
nuno.otero@dsi.uminho.pt

Nelson Zagalo
engageLab / Dept. Comm. Sciences, University of Minho
Guimarães, Portugal
nzagalo@ics.uminho.pt

Manuel João Ferreira
engageLab / Dep. Industrial Elec. University of Minho
Guimarães, Portugal
mjf@dei.uminho.pt

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.
ACM 978-1-4503-0268-5/11/05.

Abstract
This work aims to address the following question: is it possible to infer the users' expectations regarding task difficulty by watching them just before the actual start?

We present a study where people acting as evaluators determined users' expectations based on non-linguistic social signals in a 20 seconds video clip. The evaluations were performed using a five-point scale and the average error of the evaluations was of one point. Preliminary results suggest what type of signals was used by the evaluators to determine the users' expected difficulty with the task.

Keywords
Social Signals, thin slices, behavior

ACM Classification Keywords
H.5.0 Information interfaces and presentation: General.

General Terms
Human Factors, Experimentation.

Introduction
Imagine yourself in a supermarket with a recently installed self-checkout cashier machine, watching the following scene: a shopper walks by the self-payment

lane, stops, stares at the cashier machine from some distance, looks to the screen... pauses... approaches closer... reads the welcoming message, gazes around the console, grabs a product from the cart, waves it in front of the machine... waits.. looks around once again... puts the product down on the belt, waits... nothing happens, frowns, looks down to the console, grabs the product in front of the red light, once again waves it a couple of time, hears the beep, looks to the screen places it down on the belt.

Most likely this description gives the observer important clues about our user and triggers a chain of more or less accurate inferences: it seems very likely that this shopper is interacting for the first time with this machine, he/she is meeting some difficulties, and would most likely appreciate some help. To reach that conclusion we did not identify any information about the user's experience with technology, or background. All these inferences are based on the signals leaked by the user, namely the hesitation approaching the system, the pauses, the pattern of gaze and the facial expression. All these are considered social signals, the expression of one's attitude towards social interactions and interplay, manifested through a variety of non-verbal behavioral language including body postures, gestures, vocal outbursts and facial expressions [7]. Humans are good at reading these signals, our communication within a social group relies on that ability [1; 5]. Computers on the other hand are completely clueless about those social signals, missing what can constitute very relevant information on the user state, attitude and perception of a system.

Social Signals

Research has shown that the non-verbal part of communication has, in many situations, as much (or even more) effect on the human interaction than the expressed verbal messages [2: 43; 3]. The formation of social perceptions depends mainly on it [7: 1062] People convey verbal and non-verbal messages to express attitudes and emotions when they intend to do so, but, especially on the non-verbal level, they are often unaware of how much information they are leaking [2; 5; 6]. This process is so natural to the human being that, even when we are not interacting with others, we tend to repeat these non-verbal behaviors [5].

The term Social Signal Processing has been used to describe the seminal work by Pentland and his research group. Their work shows the ability, in specific social interactions, to predict the outcome of a situation or determine a person's role in a social setting based on the social signals exhibited. For example, in one application, they could predict the result of employment negotiations based on such speech features [3]. They also refer to positive results concerning the prediction of other conversational outcomes, such as professional competence, criminal conviction, divorce, or speed dating, with an accuracy of up to 70% [6]. The source of these predictions are thin slices of behavior, a short recording of the interlocutors' behavior that is sufficient to predict the outcome of a situation.

The developments in recent years with the inclusion of a variety of sensing modalities (and in particular computer vision) in HCI, with notorious popularity in games consoles, opens an opportunity for the interface to step-up and become more aware of the users.

The work discussed in this paper builds on those results and hypothesizes that the dynamics of social signals can also be valid for revealing important aspects of human-computer interaction. A system with built-in heuristics has the potential to be able to determine users´ quality of interaction or even predict users´ problematic interactions just by watching user behavior. Studying if and how others can assess or predict about the quality of human-computer interaction from the observation of the user' behavior dynamics would allow for the development of systems oriented towards those behavioral features that are most relevant.

The main question driving our research efforts is: given a particular set of heuristics derived from the users' behavior, is it possible to make predictions regarding the user's level of expertise, the quality of an interaction or the success or failure of the interaction? The work discussed here, however, focuses on the users' level of confidence towards a task.

Study Design

The experiment described aims to investigate if, even before the user engages with the task, there are relevant social signals that can reveal the users' expectation towards the difficulty of a task. Our own expectations are that such social signals include, but are not limited to, hesitation /pauses, body postures, gestures and facial expressions.

The chosen interaction context was using a photocopier. Five participants were asked to perform three different tasks on a photocopier, each task having a distinct level of difficulty: make a single page copy (easy), make a front and back copy (intermediate), make a front and back copy with two pages per side (hard).

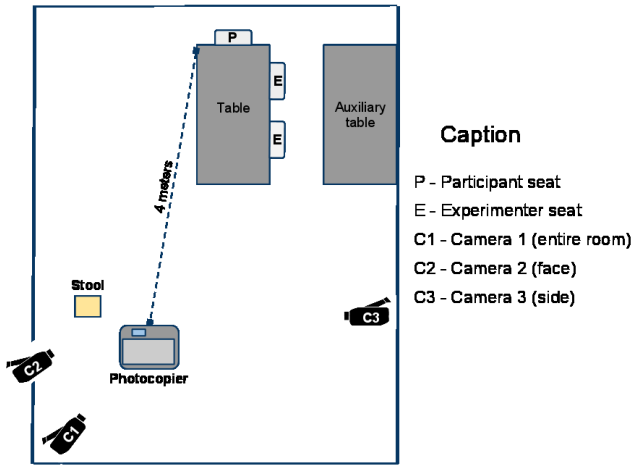

figure 1. Layout of the experiment setup.

The order of the tasks each participant performed was assigned randomly and was not the same for all. All participants had different degrees of experience in using photocopiers. The whole episode was filmed with the consent of the participants. The cameras were positioned in such way so the participants could not easily spot them, despite being aware of their presence.

Procedure

At the beginning of the trial, both the participants and two experimenters sat at the table (see figure 1). This particular location was the start and finish point of each task. Each trial was performed individually.

The participants would then be instructed on the different tasks assigned and read, one by one, aloud. After this initial step, the participants filled a form indicating the expected level of difficulty. Assessing this beforehand ensured us we could determine users' expectation without being affected by how well the task was actually performed. Then, when ready, the participants stood up and advanced towards the copy machine to perform the task. Participants were not given any instruction on how to actually perform the task, merely being indicated the desired end result. After performing the task, successfully or not, the participant would return to this initial point of departure and filled a new form evaluating the actual level of difficulty experienced with the task. The level of difficulty, both before and after performing each task, was quantified using a five-point Likert item based style scale, ranging from 1 (Easy) to 5 (Hard).

Data Analysis

We first wanted to make sure the difficulty level assigned to each task corresponded to the actual difficulty level experienced by the participants. By analyzing the participants' answers about the difficulty experienced performing each task we could determine that the easy, intermediate and hard tasks were perceived as such by the participants.

The video data was then segmented into clips, where each clip included the photocopier approach time and task preparation per task per each user. Since there were five participants performing three tasks each, there were fifteen of these videos in all. These clips were shot using camera 1 (fig. 1) which covered the movement of the user approaching and standing in front of the photocopier. Photocopier approach time measured the time between the participant getting up and arriving at the photocopier. Task preparation time measured the time between the participant arriving at the photocopier and the moment the task is initiated, *i.e.*, the participant begins interacting with the photocopier's setup menu or buttons. In average these video clips were about 20 seconds long. Our analysis showed that none of these times appeared to be influenced by the task's expected level of difficulty, since time differences across all tasks were minimal, no matter whom the participant was or what task, or task order, was being performed.

Ten other participants, from here on named as evaluators, viewed those clips through a private online webpage. For each video, evaluators were asked to indicate what they thought was the level of difficulty the participant expected to have before performing the task. They were not aware of what was the actual task being performed or what was its difficulty level. Here, the same five-point Likert item based scale, ranging from 1 (Easy) to 5 (Hard), was used. Evaluators were also encouraged to comment and justify their responses.

Evaluators' answers were then compared to those given by the participants and the following analysis was carried out: how accurately could evaluators predict the level of difficulty the participant expected to have?

For this purpose, we calculated the absolute difference between the level of difficulty each participant expected to have and the level of difficulty predicted by each evaluator. A lower absolute difference means a greater accuracy of the evaluator. Next, and for each evaluator, we calculated the average of these absolute

differences. These can be seen on the second column on table 1. The lower the average, the more accurate the evaluator is. You will notice nearly all of them are below 1.50 with the overall average being 1.23. This could indicate evaluators weren't that far off the participants' responses. We then took this analysis a step further by determining just how big a margin of error was needed before evaluators correctly predicted the level of difficulty of most of the tasks. That information can also be seen on table 1.

Evaluator	Average absolute difference	Number of tasks correctly evaluated with...	
		No error	Error <= 1
1	1,47	3	8
2	1,00	4	11
3	1,00	5	11
4	1,20	5	9
5	1,07	5	11
6	1,07	4	10
7	1,40	0	10
8	1,53	4	9
9	1,13	2	11
10	1,47	3	8
Average	1,23	3,5	9,8
Correctly evaluated tasks		23%	65%

table 1. Number of tasks where the expected task difficulty was evaluated with an error up to 1 point in a scale of 5 points.

The results presented in the "no error" column of table 1 shows that there were not many (average of 23%) tasks for which the evaluators managed to correctly predict the level of difficulty. The highest mark obtained was by three evaluators that got five tasks right. However, if we increase the margin of error by one level, an average of 65% of tasks were correctly predicted, and six evaluators got at least ten tasks right (⅔ of the tasks).

This seems to indicate that those short seconds of video were enough for most evaluators to make fairly accurate predictions about the level of difficulty the participant expected to have.

What Social Signals Did Evaluators Use?

So, what did the evaluators observe that enabled them to make their predictions or, in other words, what social signals did evaluators use? For this we turned to their comments for each evaluation. Most of these comments were either vague or ambiguous, with "hesitation" being by far the most common word used to describe a participant's behavior whenever the level of difficulty was intermediate or above it. Still, some evaluators managed to be more concise and pointed out factors like:

- Lack of fluidity - pauses between movements and unnecessary repetition of movements;
- Self-touching - touching one's own body parts or clothes;
- Prolonged staring at the photocopier's setup menu;
- Palm-up - opened palm raised (one evaluator even mentioned this as a "universal sign of uncertainty").

Also important is that most evaluators considered that the lack of any observed signals was a sign that the participant would have no difficulty performing the task, which proved right in most cases.

These results strongly suggest that evaluators had trouble explaining their predictions. This could perhaps be explained by what Gladwell, in Blink, calls "snap judgments and rapid cognition" which "take place behind a locked door", a locked door we have trouble accepting, exploring and explaining [4].

Significance for Future Work

This work is a first approach to a longer-term research goal of understanding to what extent it is possible to develop a system that could automate these judgments based on thin slices of behavioral observations. From the above indications on the social signals used, it is not clear cut how the evaluations performed and the possible underlying criteria can be easily translatable into machine detection mechanisms. There are nonetheless important indicators such as the timing, absences of movement, or break from expected patterns of movement that might reveal predictive of users' expectation towards the task.

There is a difference between describing non-verbal signals and accurately assessing what they mean. One way forward is to analyze different channels of social signaling, and to consider the contextual aspects and subtleties. However, such endeavor is not easy, and given the underlying uncertainty, other strategies might need to be utilized to entice and engage the potential users. Computers don't always get it right. However, humans' own fallibility and their own awareness of this fact have driven a long standing development of recovery processes from communicational breakdowns. Maybe, this is another line to explore instead of just assuming complex *a priori* models and their ability to make predictions.

Acknowledgements

We thank all participants who took part in this study.

References

[1] Argyle, M. 1975. *Bodily Communication*, Madison: International Universities Press, 2nd Ed

[2] Argyle, M. 1985. *The Psychology of Interpersonal Behaviour*, Penguin Books, Middlesex, USA, 1985 (4th Ed).

[3] Curhan J., A. Pentland. 2007. Thin slices of negotiation: predicting outcomes from conversations dynamics within the first 5 minutes. *Journal of Applied Psychology*, 92 (3):802-811. 2007.

[4] Gladwell, M. 2005. *Blink: The Power of Thinking Without Thinking*, Little, Brown and Company.

[5] Knapp, M. L., and Hall, J. A. 1997. *Nonverbal communication in human interaction* (4. ed.). Fort Worth, Tex.: Harcourt Brace College Pub.

[6] Pentland, A., 2008: *Honest Signals: how they shape our world*, MIT Press.

[7] Vinciarelli, A., M. Pantic, H. Bourlard, and A. Pentland. 2008. Social Signal Processing: State-of-the-Art and Future Perspectives of an Emerging Domains. *Proc. 16th int. conf. on Multimedia*, October 26-31, Vancouver, British Columbia, Canada.