Analyzing non-verbal cues in Usability Evaluation Tests

Pradeep Yammiyavar[1], Torkil Clemmensen[2] and Jyoti Kumar[1],

[1] Indian Institute of Technology, Guwahati, India, {jyoti.k, pradeep}@iitg.ernet.in
[2] Department of Informatics, CBS, Copenhagen, Denmark, tc.inf@cbs.dk

Abstract. Verbal data is the primary focus for analysis in the prevalent Usability evaluations like in 'Think Aloud Method'. This study involves 18 cross cultural TA tests and it was found that users use gestures profoundly to communicate their mental activities. It was observed that hand gestures are attempts to communicate abstract feelings as well as to quantify, to simplify a complex expression & refer to fuzzy thoughts. 10 further TA tests, with close up cameras for capture of facial expressions yielded gestures of affect states of surprise, satisfaction, confusion, deep thinking, frustration and boredom being experienced by the user. Most importantly, the users were either verbally silent or were using words seemingly incongruent to verbalisation. Observing that there is rich meaning in gestures, this paper argues for gestures as additional data sources in TA analysis.

Keywords: Think Aloud Test, Gesture Analysis.

1 Introduction

Verbal protocol analysis has become an accepted tool for usability evaluation in HCI field. Think aloud (TA) as a method of understanding the cognitive processes [1,2,3] of the user is being used extensively in both academic research and industry applications. Inadequacies of verbal expression and mismatch of verbal capacity and fluency to the speed and complexity of thought processes have also been reported [4,5]. Non verbal cues like gestures, eye movements and tonal variations are observed in the user's attempt to express what is going on in his/her mind during the task fulfillment in the think aloud usability tests [6]. Each gesture or movement can be a valuable key to recognize emotion a person may be feeling at a time. What people say is not always what they mean or are feeling [7]. According to Korchin [8] a gesture can seem as an intentional act of communication. Gestures along with bodily movements, postures, gait, facial expression and non verbal speech patterns can unintentionally yield information. Way back in 1968 Mahal [9] found that personally meaningful gestures reappeared periodically during interviews. Some movements had the same meaning and occurred simultaneously with verbal activity.

In the conducted TA tests, the selected sample users sat before a computer in a standard usability testing setup to fulfill a task predetermined by the tester/ testing team. The subjects were asked to Think Aloud while he/she performed the task/tasks.

The users were expected to be or get acquainted with the thinking aloud method as he/ she received instructions and performed a mock test. The tests were done in two stages. Having observed in stage one, the significance of gestures, stage two experiment was conceived to study the phenomena deeper. The interaction of the facilitators and users were recorded in the first stage of the experiment whereas both the interaction and the facial expressions of the users were captured in the second stage done after a gap of 9 months. 18 facilitator-user pairs of which 6 pairs consisted of facilitators from European origin and users from Indian origin were involved in the first stage of the tests whereas 10 facilitator-user pairs from the same culture took part in the TA tests in the second stage. The task in the first stage was to make an invitation card for an Indian wedding whereas the task in the second stage was to select a place of choice for picnic with friends in national tourism websites of three countries, namely, India, China and Denmark. The Recorded videos in the first part were coded for most occurring gestures and the frequency and intent of each gesture as understood by same culture two interpreters was registered. In the second stage images of gestures form the videos were extracted and shown to the acquaintances of the users for their interpretation. Finally a list of identified important gestures has been made and is being reported in this paper.

The use of non verbal cues to understand the cognitive processes of the users more reliably in the usability tests is being suggested in light of the findings from the experiment. Non verbal cues (NVC) were found in user's recorded behavior occurring repeatedly over several think aloud tests. The number of gestures was found to increase when the user was groping for words or had no words to describe a certain mental state like frustration, recall etc.

2 Method

The think aloud usability tests were conducted in two stages, in first stage, interaction behaviors and gestures of the facilitators and the users were observed at a gross level and in the second stage close up facial expressions of users were focused at. The second test was conceived after the findings from the first stage suggested the powerful role gestures were playing in the user's communications during TA.

2.1 Stage One : Think Aloud Tests with interaction analysis setting

In this stage, 18 think aloud usability tests of a task on a software known to the users with seeded errors for the task were conducted with different user - evaluator pairs. The pairs varied in hierarchy, age, gender and country of origin. 6 evaluators; 3 bachelor degree Indian students (age group 19-21 yrs); 2 European academicians and 1 Indian academician (age group 35-50yrs), each having evaluated 3 different users for the given task using think aloud method were video recorded. The task was to design a wedding card invitation for an Indian marriage. The camera was placed so as to record the interaction of the facilitator-user pair.

First, the recorded videos were qualitatively analysed by replaying them several times for identifying relevant verbal and non-verbal behaviours of users and

evaluators. Next, the non verbal behaviours were identified and were checked for their frequency. For instance, in table 4.1, subject D1, the user was raising his hand often to show something he intends to say about the images being used in the wedding invitation card making. These often repeated gestures were then registered and images were extracted.

Two independent reviewers (one in late '40s, the other in 20's, both male) were asked for their interpretation of the images. It was difficult to make meaning out of the hand gestures initially by just looking at the images. Hence the movie clips were made and then were analyzed. The mutually agreed interpretations were registered. Some of the images are presented in Table 1.

2.2 Stage Two: Think Aloud Tests with facial expression recording setting

Ten facilitator-user pairs where all were acquaintances for last three years, age group 20-23 years, nine male-male pairs, one female-female pair, did the Think Aloud test for website exploration of three countries, looking for places to visit with their friends on the national tourism websites of China, India and Denmark. The facilitators were trained during a course in usability on how to conduct TA tests. A scenario for the task was presented and then the task was introduced to the users. The users were given an approximate time of 45 minutes for all the three websites, with a minimum time limit of 10 minute per website and a maximum of 20 minutes. At the end of each test a qualitative interview regarding the interactions was conducted wherein questions regarding the interaction, TA behaviour, the satisfaction level and impact of this method on their task fulfillment were asked and then developed upon to get more insights. A close up camera captured the facial expressions of the users along with the think aloud behavior. The verbal data was recorded along with the screen capture of the activities. The recorded video was coded for the gestures. The most repeating facial expressions were registered. Also strategically most important gestures involving facial expressions were registered. Images of these identified gestures/ facial expressions were extracted from the video.

The extracted images from stage two were shown to a) the 14 acquaintances of the subject, who had known him/her for more than 3 years b) strangers to the subject but from the same age group and culture c) the user himself/herself. The subject was shown the images after 3 days of the conclusion of the tests. Because the facilitator was an acquaintance himself/herself, his/her views were also taken along with separately. The subjects, acquaintances and the strangers all were asked the same question, " What do you think, the person's expression is." All the acquaintances had undergone the tests and hence they knew the context of the images. Three reviewers who had not known the subjects (age group 20-35) were asked for their understanding of the expressions without telling them the context.

3   Results

In stage one, 14 out of 18 videos (4 were discarded for technical quality reasons) have been analyzed for gesture interpretation in form of images and clips. Images and clips

were shown to the independent reviewers from the same culture. Presented in Table 1 are some of the most occurring and strategically most relevant images of the gestures, for example. The Think Aloud behavior along with the gestures is also presented side by side. The mutually agreed interpretation by the reviewers is also presented in the last column consisting of the researchers' inferences from the gestures.

Table 1. Most occurring and strategically relevant images of the gestures
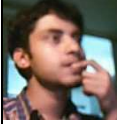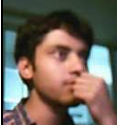
Case	Gesture Image	Behavior	Interpretation of gesture	Reason for the gesture
D1		"…the images on screen are small…"	Hand Gesture used to focus on the size of images and show by gesture that they are this much small.	The amount of smallness and resulting uneasiness has to be expressed..
D2		"…these Westernis ed Kind of …."	The invitation card has feel of western culture…the hands attempt to communicate what is difficult by words- namely western orientation.	Subconscious attempt to communicate his feel of western, which is very abstract
D3		"…it starts with a paragraph …"	The staticness of the paragraph over images or free flowing text is being intended in the card design	Communicating the abstract feeling of the staticness
D4		"…the bold letters.."	In the card design, he is thinking about the impact of the bold letters and is building his words to communicate the feeling.	Thinking and drawing attention to being surprised with the boldness effect.
D6		"…. With the figure…."	Explanation for the part of the figure that is being referred … the curvy movement of the hand shows the intended feature of the image that makes him like it	To simplify the composites of image and show the relevant element of the image that is in the context
D7		"…in Indian tradition, things like rangoli…"	Explaining to a western facilitator an Indian concept related to cards and attempts to show through hand movement	Introducing a concept (rangoli) which is alien to the facilitator hence putting all means of communication at use to make oneself understood.

D8		"..this much width.."	The card needs to be this much wide	It is difficult to quantify, state in words, the amount
D9		"… the background is…hmm…"	Thinking.	Thinking. Looking for a word

In stage two, 8 of the 10 videos (two were discarded for technical quality reasons) were identified for the image extraction of gestures. The most occurring gesture's images were extracted and shown to the a,b and c groups as mentioned above. From the pool of interpretations, common words, only differing in linguistic labels, were grouped and identified in one of the selected words. These words have been presented in bold letters in Table 2 along with the identified images. Other interpretations, not elicited commonly by most reviewers have been tabulated in normal fonts.

Table 2. Interpretations of the facial expressions of the test users

Image	Facial expression	Interpretations			Users TA Behaviour
		A (By acquaintances)	**B** (By Non Acquaintances)	**C** (By User)	
S1.1		**Fascinated**	Observing, Searching, expecting, Curiosity	Fascinated	Silence
S1.2		"Oh, I've got it", **Deciding**	Thinking, Deciding, Waiting, About to express something	Scrutinizing	Silence
S1.3		**Frustration**	Only Looking at something, not thinking much, Something went wrong, Thinking	"not able to get what I want"	Silence
S1.4		**Happy**	Seen some Familiar 'thing' and has some views, Expected result has come,	"Got what I wanted"	" Oh!, it was this" regarding an activity on the

S1.5		**Astonished**	Astonished, Something wrong, Dislike for something	Can't remember	Silence
S1.6		"Oh, What has happened, afraid, surprised	**Astonished**, Something went terribly wrong, surprised	I was bored, was yawning	Silence
S2.1		**Excitement**, Something new, enthusiastic	**Happy**	Excitement	Silence
S2.2		**Searching**	**Thinking**, Seriousness	Was Thinking	Silence
S3.1		Thinking, Bored, Load on mind, **Judging**	Observing, **Thinking**, watching seriously	Thinking	Silence
S3.2		**Confused**	**Deciding,** About to say something	Thinking	Silence
S3.3		Have found some way	**Observing,** thinking	Something Interesting	Silence
S3.4		Shy of saying something	**Thinking**	Reading and thinking	Silence
S4.1		Frustrated, **Tense**, Perplexed, in pain due to something, burdened.	**Confused,** Is forced to make decision, Tense, natural	Tired	Inaudible murmur
S4.2		**Observing and Thinking**, Determined after thought	**Worried,** natural	Observing a particular detail on the screen	Silence

S4.3		**Happy**, Got what was being looked for	**Happy**	Looking at a picture on the screen	Silence
S5.1		**Surprised**, 'what is this'	Thinking, Astonished, curious	Was observing	Silence
S5.2		**Inquisitive**, 'what is this'	**Thinking**, 'How can it be'	Observing	Silence
S5.3		About to speak	Observing, **Thinking**	First time looking at something	Silence
S6.1		Found some known thing	**Happy**	Normal look	Silence
S6.2		What does this mean?	**Observing**	Thinking	Silence
S7.1		**Thinking**	Thinking seriously, trying to make some opinion, observe	Thinking	Silence
S7.2		**Suspicious**	Little Confused, trying to make opinion, observe	Observing	Silence
S7.3		Correlating something	Quite Worried, Confused, observe	Observing	Silence
S7.3		Surprised, **Inquisitive**	Seen a ray of hope	Annoyed	Silence

| S8.1 | | Bored, doubtful | Doesn't agree to what he is watching, observe | Observing | Silence |
| S8.2 | | Concentrating | Watching, normal | Thinking | Silence |

Table 3. Common facial gestures among eight test users in usability tests

Subject	Most Recognizable facial Gestures Elicited
S1	Surprised, Deep thinking and Bored
S2	Happy and thinking
S3	Confused
S4	Happy, Worried and Puzzled
S5	Frustrated and Happy
S6	Surprised and Inquisitive
S7	Thinking and Happy
S8	Concentrating

4 Discussion and conclusion

We have five observations from the analysis of gestures and facial expressions of test users during think aloud usability tests:

1. Subjects varied in kind of gestures made though a few gestures were common to all, namely, happiness, boredom, frustration and deep thinking (Table 3). Also the number of gestures made were different in different users. For instance the subject S1 and S7 had shown more gestures as compared to S3 and S8. Also the think aloud behaviour was more in duration and came naturally to some of the subjects without they being aware of it, like S1 and S9 while it was very little in the case of S3 and S7. The individual differences in the TA behaviour and the gestures have been found to compensate for the verbalisations in the usability tests in the case of S7 where more gestures and less TA behaviour was observed and accentuate in case of S1 where both were in greater magnitude comparatively making the subject more easy to analyse for the satisfaction in interaction with the website. In cases of other subjects (S2-S6 and S8-S9) the gestures and verbal data enrich each other, sometimes complementing (when both gesture and think aloud behaviour is there) and on other occasions supplementing (when the subject is silent, but there is/are gesture/s or vice- versa). In case of subject S3 the gestures become crucial for there is less of TA and less of gestures. For thinking aloud doesn't come naturally to the user and by less gestures here we only mean less no. of gestures, for the person is

always in some position and the *face can still be the index of mind*, the only issue is of the agreeing upon the meaning of the expression.

2. We observed that 5 out of 10 subjects reported that the TA behaviour interfered with that part of the task which involved thinking and deciding from several choices, for example, which hotel to book, depending on the aesthetics, cost, location etc. Our inference in this case is that the cognitively loaded tasks demanding more of mental resources and hence making the Think aloud behaviour an interference with the task itself. This has been an oft reported phenomena related to think Aloud and these findings only substantiate it. Therefore it is posited that gestures are brought into play to reduce the cognitive load.

3. Subjects (6 out of 10) reported the exhausting effect of think aloud. They reported that the TA activity made them tired and 2 of them felt even hungry at the end of the tests. Our inference here is that the extra effort spent in thinking aloud, specially when a facilitator keeps on nagging to keep thinking aloud may also have physiological effects on the body of the subject, in terms of stress, strain and fatigue. And it is highly probable specially in cultures like India that these physiological loads may be expressed more through gestures rather than verbally, similar inferences have been reported by other researchers [7].

4. It was sufficient in the facial expression inference case to just show the image (except for the case of yawning) while the hand gesture interpretation required longer movie clips of the gestures in order to make judgement about them. It seems that the hand gestures involved spatial and time domain activity and were used to accompany the verbalisation while the facial expressions were more tacit, and affect related and were *' there on the face',* hence only images gave good results and satisfaction was reported by the interpreters in case of facial expressions.

5. The interpretations of the facial expressions shown in the images were found to give cue to almost same understanding of the gestures by the user (except for the case of yawning image, S1.6) and their acquaintances, whereas the strangers' interpretations had deviations from reports of the user. The user's and facilitator's reporting, upon showing the image was more in terms of what the person was doing at that instance, than what the expression meant. This is a positive observation towards developing cultural gesture recognition protocol, for the acquaintances are able to predict the expression in the image. The failure on the part of strangers on the other hand challenges such a possibility, as they too are from the same culture but are not able to identify the expression in the image.

To conclude, from the above series of observations and inferences it is clear that gestures carry substantial amount of information which can be tracked to enrich/ complement or supplement the verbal data obtained during the usability tests aimed at knowing the cognitive processes of the users mind. Researchers like Albert Mehrabian [11] have held that transmission of message is effective only when all the three aspects of communication – the verbal (words - 7% impact) , the vocal (intonation, pitch ,volume - 38% impact) and the visual (gestures ,postures –55% impact) are in tandem with one another. In light of the above inferences it is posited that some of the nonverbal behaviors expressed through gestures can act as clarifiers of the communication that is happening in a standard think aloud protocol situation.

5 Future Work

The gestures can further be analyzed for cross cultural interpretation of static images of gestures and movie clips of the gesture elicitation. People from other cultures can be subjected to these images and asked what they understand from it. Table 1 and Table 2 can further be detailed out with time duration and strategic locations of the gestures in the usability tests, including the role of silence. Further work can be done towards developing a culture based 'gesture lexicon' which can be used by cross cultural usability testing professionals using TA method.

References

1. Maarten W. van Someren, Yvonne F. Barnard, Jacobijn A.C. Sandberg (1994); (p. 26) The Think aloud method - A practical guide to modelling cognitive processes, Academic Press, London.
2. Nielsen .J, Clemmensen .T, Yssing .C (2002); Getting access to what goes on in people's heads? - Reflections on the think-aloud technique, Nordi CHI.
3. Ericsson.K & Simon. H (1993); Protocol Analysis – Verbal Reports as Data, MIT.
4. Boren, M. T. & Ramey, J. (2000), Thinking aloud: Reconciling theory and practice, IEEE Transactions on Professional Communication, vol. 43, no. 3, pp. 261-278.
5. Branch .J, (2000), The Trouble With Think Aloud: Generating Data Using Concurrent Verbal Protocols . Proceedings of CAIS 2000: Dimensions of a global information science.
6. Yammiyavar. P and Goel K M (2005), Emphasis on non-verbal cues for interpreting cognitive processes in protocol analysis, Proceedings Indo Dan Research Symposium on HCI, Guwahati.
7. Pease .A & Pease .B (2005) The Definitive Book of Body Language, Pease International Pty Ltd, Australia.
8. Sheldon. J Korchin; Modern Clinical Psychology; Basic Books, New York, 1999.
9. Mahl .G.F; gestures and Body movements in interviews in J.M.Shlien (Ed) Research in Psychotherapy Vol 3 , American Psychology Association, Washinton DC 1968.
10. Mehrabian, A. (1981). Silent messages: Implicit communication of emotions and attitudes (2nd ed.). Wadsworth, Belmont, California.