Evaluating Interfaces: Thinking Aloud

COMP2044: Human-Computer Interaction (2024-2025)

Matthew Pike

Overview

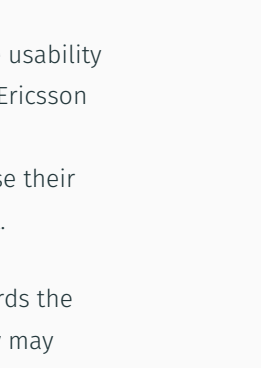
Objectives for today

- Using Think-Aloud Protocols (TAP) as a passive evaluation method.
- Cooperative Evaluation as an active evaluation method.
- Overview of Coursework 2.

Think-Aloud Protocol (TAP)

What is TAP? (Ericsson & Simon, 1980)

- An extremely popular and simple usability evaluation method proposed by Ericsson and Simon in 1980.
- Participants are asked to verbalise their thoughts while performing a task.
  - That's it!
- The evaluator observes and records the participant's verbalisations. They may need to prompt the participant to keep talking.
- The method has been shown not to impact participant mental workload or task performance (Pike et al., 2014).



Session: Brain Computer Interfaces CHI 2014, One of a CHind, Toronto, ON, Canada

Measuring the Effect of Think Aloud Protocols on Workload using fNIRS

Matthew Pike[1], Horia A. Maior[1,2], Martin Porcheron[3], Sarah Sharples[3] and Max L. Wilson[1]
[1]Mixed Reality Lab, University of Nottingham, Nottingham, UK
[2]Horizon Doctoral Training Centre, University of Nottingham, Nottingham, UK
[3]Human Factors Research Group, University of Nottingham, Nottingham, UK
{psxmp6, psxhama, psxmp9, sarah.sharples, max.wilson}@nottingham.ac.uk

ABSTRACT
The Think Aloud Protocol (TAP) is a verbalisation technique widely employed in HCI user studies to give insights into user experience, yet little work has explored the impact that TAPs have on task performance. Functional Near-Infrared Spectroscopy (fNIRS) is a brain sensing technology that offers the potential to provide continuous, detailed insight into brain activity, enabling an objective view of cognitive processes during complex tasks. Participants were asked to perform a mathematical task under 4 conditions: nonsense verbalisations, passive concurrent think aloud protocol, invasive concurrent think aloud protocol, and a baseline of silence. Subjective ratings and performance measures were collected during the study. Our results provide a novel view into the effect that different forms of verbalisation have on workload during tasks. Further, the results provide a means for estimating the effect of spoken artefacts when measuring workload, which is another step towards our goal of proactively involving fNIRS analysis in ecologically valid user studies.

ACM Classification Keywords
H.5.2 [Information interfaces and presentation]: User Interfaces – Evaluation/methodology.

Author Keywords
functional near-infrared spectroscopy; fNIRS; BCI; human cognition; Think Aloud Protocol; HCI

INTRODUCTION
The Think-Aloud Protocol (TAP) is a widely used research method [2,6], utilised in a variety of research fields including Human Computer Interaction. Since TAP will use resources from short term working memory, it is fair to assume that the inclusion of spoken protocols will potentially affect cognitive processes due to use of available resources. Consequently, TAPs may affect performance in tasks, and also measures of workload during studies.

To analyse the potential impact that a TAP may have on an individual, we use a direct measure through the brain monitoring technology, Functional Near-Infrared Spectroscopy (fNIRS). fNIRS has received recent focus in HCI research for its amenability for more ecologically valid study conditions [17, 33]. While some brain sensing techniques like functional Magnetic Resonance Imaging (fMRI) require minimal or no movement from users, fNIRS can can be used while seated naturally at a computer [33]. Further, because fNIRS measures blood oxygenation and deoxygenation rather than electrical signals like Electroencephalography (EEG), fNIRS permits more natural movements associated with using a computer without being subject to significant artefacts in the data. Although the suggestion is that fNIRS can be used more easily within natural, ecologically valid user study conditions, current research is still limited to performing controlled simple Working Memory tasks (e.g. [33, 30]).

In the context of HCI, TAP is typically used as an evaluation method to elicit insights into participants thoughts and strategies during usability and user studies. TAP however, has also been used in other settings, such as cognitive psychology and social sciences [7], to understand phenomena such as user mental models, expertise, and problem solving. As well as being a core part of user studies, verbalisations are also closely related with Working Memory, as both the interpretation of words in the task and the integration of thoughts involve the phonological loop [18]. Consequently, to integrate fNIRS measurement within a typical user study that might involve a TAP, we have to be aware of how one will affect the other. There are various forms of TAP, including retrospective, which occurs after a task has been completed, and concurrent, which occurs during a task. Of concurrent forms of TAPs, there is both invasive, which involves directly questioning participants, and passive, which only encourages participants to maintain verbalisations about their thoughts and actions. Because fNIRS measurements are taken during tasks, this paper focuses on concurrent TAPs.

In the following sections, we first review related work on TAPs, Working Memory and mental workload, fNIRS sensing and other technologies. The paper continues by describing a user study examining the impact of a) nonsense verbalisations, b) passive concurrent think aloud, and c) invasive concurrent think aloud, compared to a baseline of silent non-verbal working memory. We then present the results of the study, discuss the findings in terms of what we can learn

Figure 1: An example of HCI research using and validating TAP (Pike et al., 2014).

In-Class Example: Performing TAP on the University Website

1. Recruit a representative user as a participant.
    - I need a volunteer!
2. Obtain informed consent from the participant.
3. Give the participant a representative task.
    - “Find who the Provost of the University is.”
4. Ask the participant to think aloud while performing the task.
    - “As you are doing this task, please say out loud what you are thinking.”
    - Occasionally, they may require gentle reminders or cues to keep talking.

Activity: Pros and Cons of TAP

Pros

- To do in class.

Cons

- To do in class.

Cooperative Evaluation

What is Cooperative Evaluation? (Monk et al., 1993)

- Employs TAP, but with a twist.
  - The evaluator also participates actively in the discussion.
- Passive vs Active TAP:
  - In Ericsson and Simon's original TAP, the evaluator is a passive observer. They only prompt the participant to keep talking.
  - In Cooperative Evaluation, the evaluator is an active participant in the discussion.
    - They may ask questions, provide clarifications, etc.
    - The evaluator's role is to facilitate the discussion, not to dominate it.
    - Care is needed to avoid biasing the participants' responses with leading prompts/questions.

In-Class Example: Performing Cooperative Evaluation on the University Website

1. Recruit a representative user as a participant.
    - I need another volunteer!
2. Obtain informed consent from the participant.
3. Give the participant a representative task.
    - “Find the academic calendar for the current semester”.
4. Ask the participant to think aloud while performing the task.
    - “As you are doing this task, please say out loud what you are thinking.”
    - During the task, the evaluator should ask questions to learn more about the participant’s thought process. Examples:
        - “Why did you click on that link?”
        - “What is your strategy for finding the information?”
        - “You seem to be struggling with this task. Can you tell me why?”

General Considerations for Collecting Observational Data

- Qualitative or Quantitative?:
  - One is not better than the other, but they serve different purposes.
  - Consider carefully what you want to know, and how you will use the data.
- What to Record?:
  - User actions - success; errors; time taken; expressions of frustration; excitement; etc.
  - User comments - verbalisations; questions; etc.
- How to Record?:
  - Audio recording; video recording; screen recording; notes; etc.
- Ecological Validity?
  - How closely does the evaluation environment need to resemble the real-world environment?
- How do you report the data?
  - Consider the audience and the purpose of the evaluation.
  - Quantitative date: graphs, tables, etc.
  - Qualitative data: quotes, summaries, etc.

Coursework 2

Task

Overview

- Evaluate a team member's prototype from Coursework 1.
- Employ established HCI evaluation methodologies:
  - First, perform a Cognitive Walkthrough.
  - Then, conduct a Cooperative Evaluation.
- Document and present your findings.
- This exercise emulates a practical, real-world assignment.

Deliverable

- A report detailing:
  - Key findings from the Cognitive Walkthrough.
  - Cooperative Evaluation results and discuss key findings;
    - These inform - recommendations for the future.
  - Refer to design principles in your discussion!
  - References!

Process

1. Cognitive Walkthrough:
   - Each group member will perform a Cognitive Walkthrough on another group member's prototype.
   - The group will convene to discuss the findings and select one prototype for evaluation.
   - The group will decide on the selection criteria and document the process. You will need to document this process in your report.
2. Cooperative Evaluation:
   - Identify the tasks you will ask participants to perform.
   - Allocate roles/responsibilities to team members to perform during the Cooperative Evaluation.
   - Recruit participants.
3. Analyse Results:
   - Identify challenges/success in the prototype design.
   - Redesign aspects of the prototype, if necessary.
   - Document the process and findings in your report.

Logistics

Recruiting Participants

Remember, you can only recruit participants from other groups in the HCI module. You cannot recruit participants from outside the module. It's therefore essential that you plan and prepare to be ready to collect data in Week 10.

- All future workshop sessions will be dedicated to Coursework 2.
- Week 10 (w.c. 21/04/2025): the lecture and workshop will be dedicated to data collection.
  - By this point, you will need to have:
    - Completed the Cognitive Walkthrough and selected a prototype to evaluate.
    - Decided on the tasks you will ask participants to perform.
    - Prepared any necessary materials (e.g., consent forms, task descriptions, etc.).
    - Have a plan/protocol for running the Cooperative Evaluation.

Report Structure and Content

- Document the key-findings from the Cognitive Walkthrough.
  - Just the key findings, not a detailed report. This should be brief, but strongly justify why you selected the prototype for evaluation.
- Document the methodology for the Cooperative Evaluation.
  - Include the tasks you asked participants to perform.
  - Include the roles/responsibilities of each group member during the evaluation.
- Document the key-findings from the Cooperative Evaluation.
  - Discuss the key findings and how they inform recommendations for the future.
  - Refer to design principles in your discussion.
  - Include any redesigns you made to the prototype.
- Remember, we want to see evidence of your understanding of HCI principles and how you applied them in your evaluation.
  - Use references to support your arguments.

Web Articles

- Thinking Aloud: The #1 Usability Tool
  - https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/
- Think aloud study: qualitative studies
  - https://www.gov.uk/guidance/think-aloud-study-qualitative-studies
- Influences on the Uptake of Health and Well-being Apps and Curated App Portals: Think-Aloud and Interview Study
  - https://mhealth.jmir.org/2021/4/e27173
- Using moderated usability testing
  - https://www.gov.uk/service-manual/user-research/using-moderated-usability-testing

References

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*(3), 215.

Monk, A., Davenport, L., Haber, J., & Wright, P. (1993). *Improving your human-computer interface: A practical technique*. Prentice-Hall New York.

Pike, M. F., Maior, H. A., Porcheron, M., Sharples, S. C., & Wilson, M. L. (2014). Measuring the effect of think aloud protocols on workload using fNIRS. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3807–3816.