Visual Cue Streams for Multimodal Dialogue Interaction

Dimitris Koryzis, Christos V. Samaras, Eleni Makri, Vasilios Svolopoulos and Dimitris Spiliotopoulos

Abstract This work examines the visual information streams as feedback to users engaging in multimodal interaction during specific settings. It reports on the findings from complexity and frequency of information presentation paired to user acceptance. It also addresses technical design issues by examining how multiple visual streams are presented in real situations, for the specific complex use case.

Keywords Human factors · Interaction design · Multimodal interaction

1 Introduction

For learning applications, a major parameter to account for is cognitive load [1]. During the design of an instructional system, cognitive load theory (CLT) is part of the human centered design [2]. Such approaches integrate HCI principles with CLT [3, 4]. This is especially evident in web based systems and can be part of the design process [5].

D. Koryzis (✉) · V. Svolopoulos
Hellenic Parliament, Athens, Greece
e-mail: dkoryzis@parliament.gr

V. Svolopoulos
e-mail: v.svolopoulos@parliament.gr

E. Makri
Hellenic Parliament Foundation, Athens, Greece
e-mail: el.makri@parliament.gr

C.V. Samaras · D. Spiliotopoulos
Distributed Computing Systems, Institute of Computer Science,
Foundation for Research and Technology, Hellas, Heraklion, Greece
e-mail: csamaras@ics.forth.gr

D. Spiliotopoulos
e-mail: dspiliot@ics.forth.gr

J.I. Kantola et al. (eds.), *Advances in Human Factors, Business Management, Training and Education*, Advances in Intelligent Systems and Computing 498,
DOI 10.1007/978-3-319-42070-7_50

Metalogue is an EU-funded project that aims to design a dialogue system with metacognitive capabilities from natural spoken language and multimodal interaction. The goal of the approach is to use multimodal dialogue to help train metacognitive skills in educational settings. In order to do that, Metalogue uses an avatar that engages in multi-issue bargaining with the users via natural language. Specific events that are used for reflection are detected and presented to the users and tutors post-dialogue in order to meaningfully visualize the user progress. In order to design the usability evaluation, three major aspects had to be explored and taken into consideration, the multimodal dialogue, the formal training of specific skills and the overall user experience.

Currently, tutors, as an additional activity to the student formal training at school (debating course), encourage the students to watch recorded debates in order to acquire skills about strategy, presentation, political awareness, as well as social skills. The proposed system builds on the advantage of the multimodal interaction to both drive the dialogue and assess user input. It is also a means for training as well as a complex system. Effectively, the usability evaluation design for such focused system, should include all three aforementioned aspects, multimodal interaction, skill training and overall user experience. This paper presents the experimentation on the visual streams of information, especially in regards to accommodating the particulars for skill training, examining the visual paradigm of real-time multi-source information presentation.

As part of the design of the system-user interaction, this work aims to experiment on the visual cues necessary for the system feedback to the user, exploring the user interaction with the system and the response to the visual signals.

The evaluation settings involve initially student members and their tutors in the debating sessions of the Hellenic Youth Parliament using multimodal interaction, engaging in dialogue in natural language with the avatar agent.

2 Experimental Setup

Five experienced Hellenic Youth Parliament members debated in pairs while the system recorded the session (Fig. 1). An in-action feedback mechanism provided visual feedback to each participant in real time about posture. The sessions lasted 10–15 min each. During the participant interaction the system provided visual real-time feedback according to specific parameters:

- Visual cues (signs) regarding feedback about presentation/posture with variable frequency
- Same as above with fixed frequency
- Multiple visual cues regarding mixed feedback about presentation and progress with variable frequency.

Fig. 1 Pilot session: users debating with system as observer

At the time, the system itself did not present any real-time audio feedback to the users. However, the sessions were recorded and reviewed afterwards as part of the user feedback process.

The goal was to evaluate the load of information presented in real time to the user in order to achieve completeness and informativeness in real time. The user feedback was provided after setting up human-to-human dialogue sessions over a selected topic for political debate. A display and the Kinect module were used to provide feedback on human movement while information from all three aspects was presented when available. The visual notifications were scripted and provided in varying densities. The participants were asked to provide verbal feedback in between short breaks of the debating. Furthermore, they were asked to fill in an online form after the end of all sessions. Each session was adapted by the feedback from the preceding one.

After the debating sessions, the participants were debriefed on the interaction experience and system feedback, mainly on the visual cues during the debate sessions. Furthermore, an online questionnaire survey was compiled and focus group discussions were organized to collect feedback and opinions, in order to better identify the necessary features of the proposed approach. The focus groups involved debate students, politics-oriented tutors of the students, parliamentary officers, policy analysts as scientific advisors, interaction and content designers.

The type and clarity of the visual indicators for the interaction progress were discussed based on the results, while the user acceptance scores for combined complexity and frequency of visual cues were collected from the user feedback session.

3 Results and Evaluation

In order to accurately account for the human experience with multiple visual input over the course of the speech interaction between two participants, the duration of the debating session time was segmented into five second intervals. The times when the visual input was present was addressed within those intervals. For the first session, the visual cue was single and was updated every 10s, that is every 2 segment intervals. The duration of the cues on the screen was 3–5s, which was based on earlier experimentation regarding the complexity of the selected visual itself and the familiarity of the participants. For three sessions, visual cues were presented in variable frequency. For session 1, new visuals were triggered for 50 % of the time segments, since the frequency was fixed. For the variable frequency sessions, the visuals were triggered for 49, 59 and 56 % of the time segments, respectively (Fig. 2).

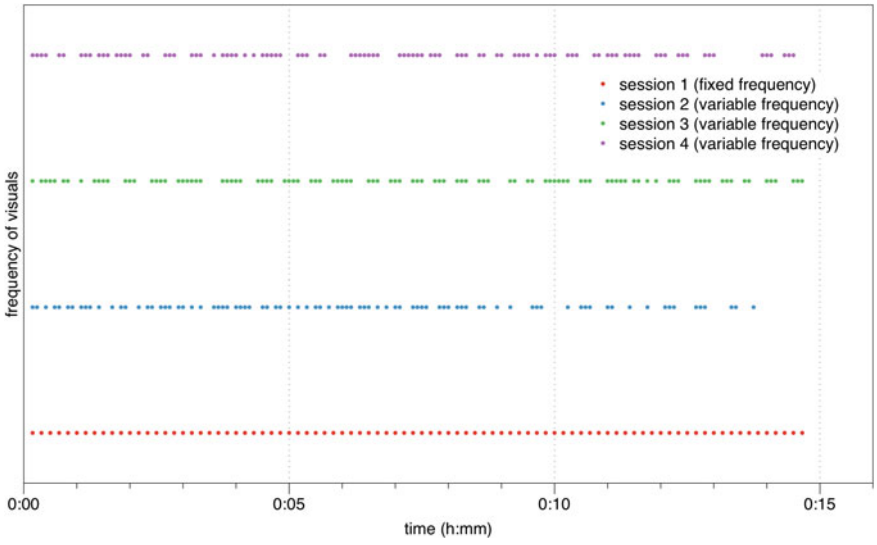

Fig. 2 Session frequency of visuals

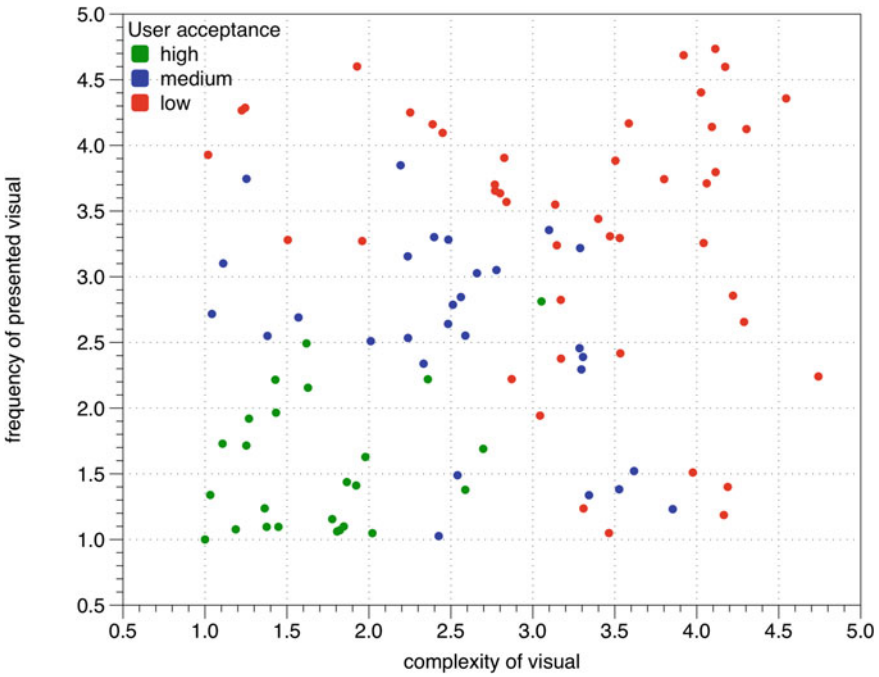

Fig. 3 Perceived complexity and frequency of visuals

The last session involved the use of multiple streams (up to three) of visual cues that were designed to closely resemble the expectations of the final design of the system.

The participants evaluated the perceived level of complexity of the visuals and the frequency that they were presented in a 1–5 Likert scale.

Additionally, their subjective feedback on the user acceptance for each instance (complexity/frequency combination) based on their ability to understand and utilize the input. As expected, the user acceptance was dependent to the complexity and frequency, rendering only 25 % of the visual signals as acceptable by the users (Fig. 3). It was also evident that complexity was easier to compensate for by the users, while the frequency (defined also as speed at which the information was presented) was harder to follow.

The aim of the last session was to observe how such approach could be designed to optimally present the information to the users while not overloading them with information. It served as a first indicator for the potential solution to managing the extraneous cognitive load. One of the three visual streams as active for 50 % of the total time, while the other two for 37 and 38 %, respectively. Information from as least two streams was concurrently present for 31 % of the time while all three streams overlapped 8 times (one for a quite significant amount of time), accounting for 8 % of the session duration (Fig. 4).

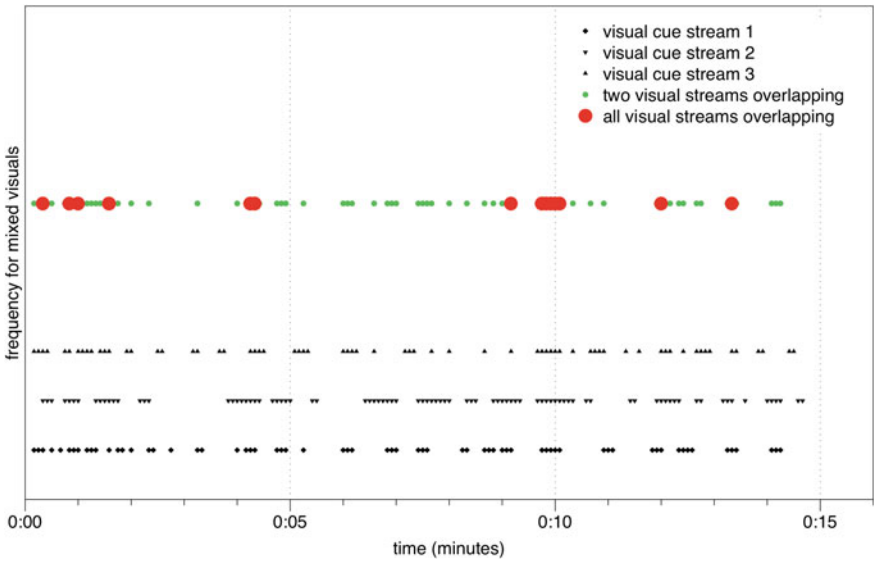

Fig. 4 The experiment location set up

4 Conclusion

The participant feedback, along with the session logs were analyzed for extraction of the findings from the subjective user feedback as well verification by the log data. It was clearly shown that a participant engaged in high-level complex interaction with another human, has very limited attention span to visual stimuli, even when that is purposed to assist them for the interaction. The need for post-processed information from infrequent high impact semantic visual notification was also clear. Additionally, focused feedback on one aspect of training was preferred to the fused approach that is also used on real life tutoring, for real time visual feedback, although that view was reversed for summative feedback after session, when the participants may reflect on their activity.

Further work on this subject is to examine the dynamics of the cognitive load on the participants under the conditions examined in this work based on the relevant related works [6]. This analysis is expected to provide insight on how the cognitive load is managed in order to optimize learning [7, 8].

Acknowledgments This work was conducted within the scope of the Metalogue project and was partially funded by the European Commission under the grant agreement number 611073. The authors would like to thank the Youth Parliament students that participated in the pilot experimentation and provided valuable feedback.

References

1. Nguyen, F., Clark, R.C.: Efficiency in e-learning: proven instructional methods for faster, better, online learning. Learning Solutions (2005)
2. Oviatt, S.: Human-centered design meets cognitive load theory: designing interfaces that help people think. In: Proceedings of the 14th Annual ACM International Conference on Multimedia, pp. 871–880 (2006)
3. Hollan, J., Hutchins, E., Kirsh, D.: Distributed cognition: toward a new foundation for human-computer interaction research. ACM Trans. Comput.-Hum. Interact. (TOCHI). **7**(2), 174–196 (2000)
4. Hollender, N., Hofmann, C., Deneke, M., Schmitz, B.: Integrating cognitive load theory and concepts of human-computer interaction. Comput. Hum. Behav. **26**(6), 1278–1288 (2010)
5. Feinberg, S., Murphy, M.:Applying cognitive load theory to the design of web-based instruction. In: Proceedings of the 18th annual ACM International Conference on Computer Documentation: Technology & Teamwork, pp. 353–360 (2000)
6. Sawicka, A.: Dynamics of cognitive load theory: a model-based approach. Comput. Hum. Behav. **24**(3), 1041–1066 (2008)
7. Bannert, M.: Managing cognitive load: recent trends in cognitive load theory. Learn. Instr. **12** (1), 139–146 (2002)
8. Wouters, P., Paas, F., van Merrienboer, J.J.G.: How to optimize learning from animated models: a review of guidelines based on cognitive load. Rev. Educ. Res. **78**(3), 645–675 (2008)