Making an Online Whiteboard Ready for Multimodal Interaction: Integrating Text- and Voice Chat into Hyperchalk

Sebastian **Gombert**[1,*], Lukas **Menzel**[2], Onur **Karademir**[1], Daniele **Di Mitri**[1] and Hendrik **Drachsler**[1,2,3]

[1]*DIPF: Leibniz Institute for Research and Information in Education, Germany*
[2]*Goethe Universität Frankfurt, Germany*
[3]*Open University of the Netherlands*

Abstract

Hyperchalk is an open-source online whiteboard aimed at use cases in education and has been developed with Learning Analytics in mind. It allows users to sketch and draw in a collaborative environment. Similar to commercial whiteboard tools, this allows the implementation of various collaborative learning tasks. However, unlike commercial solutions, in the background, the tool collects trace data, which can be used to study learners' collaboration processes and calculate metrics that inform teachers about their learners' collaboration processes. Collaboration, however, always involves communication between collaborators, and analyzing it is crucial for understanding emerging learning processes. So far, this was a dead spot of Hyperchalk as the tool was limited to collecting only what happened on the whiteboard itself. In this paper, we present an updated version of the tool, which allows users to communicate directly within Hyperchalk with the help of text and voice chat functionalities that have been added. In particular, we illustrate the technical architecture that is a basis for these added features.

Keywords

Online Whiteboard, Multimodal Learning Analytics, Voice Chat, Text Chat, Computer-supported collaborative learning

1. Introduction

Hyperchalk is an open-source online whiteboard. It is intended to function as a tool for implementing dynamic, collaborative learning activities in computer-supported collaborative learning scenarios and collects rich log data about the collaboration processes to understand the learners better and support them through Learning Analytics [1]. However, so far, this data collection has ignored essential components of computer-supported collaborative learning: oral and written communication. In current synchronous online learning scenarios, communication

is mainly managed through video conferencing tools, which provide limited access to communication data. For Learning Analytics research, this can be problematic as it can be hard to fully understand the collaboration processes that occur during collaborative learning activities without analysing the actual communication of the people involved [2].

For this reason, we updated Hyperchalk with voice-over-IP and text chat functionalities. Users can now use these functionalities on all whiteboards, allowing instructors to deactivate them optionally. For implementing these features, we chose a centralized approach so that all communication between learners is sent over the Hyperchalk server. This guarantees that it can be fully logged and is available for real-time or post-hoc analyses. Moreover, to allow practitioners to create interactive learning experiences with Hyperchalk, we offer APIs to integrate external bots with the tool that can send learners text- and voice messages and manipulate the board state.

2. Background

Work on computer-supported collaborative learning is often based on constructivist pedagogy. Speaking in simple terms, this school of thinking follows the notion that learners don't simply learn by transmitting and receiving learning content. Instead, they need to (re-)construct their understanding of what is to be learned from the input they receive. Following this notion, in computer-supported collaborative learning scenarios, it is often emphasized that learners communicate to help extend each other's understanding of a given topic [3, 4].

In collaborative learning scenarios, learners often take on different team roles [5]. These influence what and to which degree learners can acquire knowledge in these scenarios. Sometimes, such roles are explicitly assigned, but more often, they emerge naturally from the individual group dynamics. Dowell et al. [6] found that such emerging roles can be identified in multiparty communication employing vector-based natural language processing methodology. Menzel et al. [7] followed up on this work by identifying roles as a basis for providing learners with role-specific feedback.

Online whiteboards allow the implementation of a whole range of different collaborative activities. Like analogous whiteboards, users can collaboratively draw, sketch and illustrate in group contexts [1]. This usually goes hand-in-hand with discussions and extensive communicative processes. Consequently, computer-supported collaborative learning is often highly multimodal by nature [2], and whiteboard-based learning tasks are no exception. We suspect that analyzing both learners' communication and interactions on the whiteboard is crucial for understanding the entire learning processes that emerge during such scenarios.

Multimodal learning analytics is the field of research dedicated to studying and understanding multimodal data produced by learners during such scenarios [8]. In this context, the look at multiple modalities aims to acquire a complete picture of what happens during learning, as different aspects of learning can be expressed through different modalities. On the one hand, this involves the products learners create during learning activities, which can be multimodal by themselves. On the other hand, learning processes can be expressed through different modalities, particularly during computer-supported collaborative learning where communication plays a key role [2]. We suspect that multimodal indicators extracted from both learners' whiteboard

interactions and communication can be utilized to identify learner roles and to provide learners with individualized feedback similar to Menzel et al. [7].

3. Method

To make Hyperchalk ready for multimodal learning analytics in computer-supported collaborative learning scenarios, we extended the tool to support voice- and text chat to allow the collection of learner data from these modalities, which can then be analyzed in combination with trace data from the whiteboard itself.

We added two additional buttons to the interface to make the chat functionalities available to the users in Hyperchalk. The goal was to integrate them into the existing user interface seamlessly. For this reason, we added two respective buttons to the tool's menu bar. Figure 1 depicts a screenshot of the updated menu bar of Hyperchalk with these additional buttons.
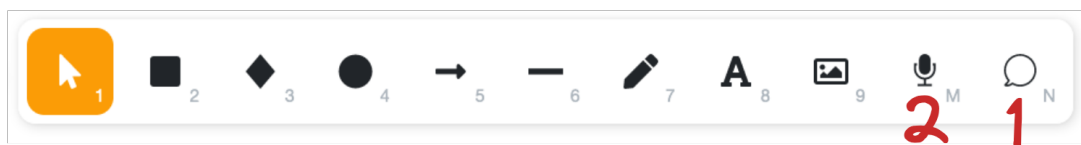

Figure 1: A screenshot of the updated menu bar with buttons for access to the new chat features added.

The first button (1) opens a chat window where users can exchange text messages. They are also provided with a simple emoji picker similar to the ones found in established messaging apps. Moreover, they can send images and files. This aims to make the user experience similar to established messengers so that users are not driven to use them instead of our integrated chat function. A link preview function is also provided with the chat.

The second button (2) allows users to turn their microphones on and off. This feature can be immediately used after granting the browser access to users' microphones. All users are automatically in the voice chat when they enter the room and can hear other participants if their microphones are turned on. This is intended to keep barriers for users to engage in communication low.

Both communication features use a centralised approach. This means all communication is sent to the server, which broadcasts it to other users in a room. The reasoning is to make all communication traceable without encountering synchronization issues that are more likely to emerge for a decentralized approach. In the following sections, we describe the technical properties of both functions.

3.1. Text Chat

For transmission of chat messages, we use a simplified version of the XMPP protocol [9], an established open text messaging standard. Messages are sent to the server, which then broadcasts them to the other clients in the session. Moreover, all chat messages are stored in the database for later analysis. The replay mode of Hyperchalk also supports playback of the chats.
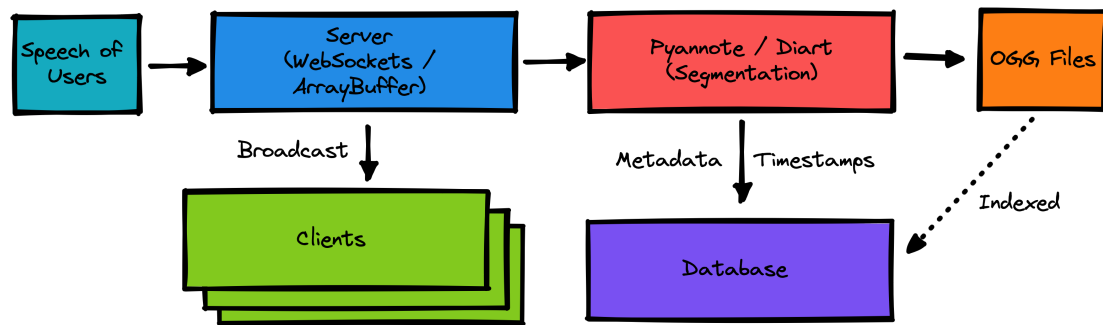
3.2. Voice Chat

Figure 2: This diagram depicts the technical architecture of the voice chat feature and the segmentation pipeline.

We use WebSockets to transmit audio messages between the server and clients. All audio messages are stored on the server, as well. For this purpose, incoming audio streams are buffered and stored with a timestamp. A constantly running background task then segments the buffered audio streams into individual speech segments using Pyannote [10, 11] and Diart [12]. The audio is encoded using the Opus codec and stored as OGG files on the server using Pedalboard [13]. An individual ID and a timestamp are generated and stored in the database for documentation for each file. Figure 2 depicts this architecture schematically.

3.3. API for External Services

On top of the chat functionalities, we offer an API for external services to interact with Hyperchalk. These external services can access respective board states and manipulate them to react dynamically and interact with users. This allows for implementing flexible, interactive learning activities in Hyperchalk through external backend services.

In this context, we also allow for the exchange of audio streams and chat messages between Hyperchalk and external services. On the one hand, this allows practitioners to develop and couple live analytics services like live transcription, affect detection or discourse analysis with the software. On the other hand, external services can use these APIs to pipe audio streams and chat messages into rooms so that it is possible to integrate voice-driven chatbots or virtual instructors that interact with the learners via Hyperchalk.

4. Conclusion

In this paper, we introduced and outlined text- and voice-chat features for Hyperchalk. With these, we aim to make the software ready for multimodal learning analytics to study the collaboration processes of users within boards. Moreover, we introduced the option to provide external services that interact with users through text and audio and react to and manipulate the board state with Hyperchalk.

References

[1] L. Menzel, S. Gombert, D. Di Mitri, H. Drachsler, Superpowers in the classroom: Hyperchalk is an online whiteboard for learning analytics data collection, in: I. Hilliger, P. J. Muñoz-Merino, T. De Laet, A. Ortega-Arranz, T. Farrell (Eds.), Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption, Springer International Publishing, Cham, 2022, pp. 463–469.

[2] S. Praharaj, M. Scheffel, H. Drachsler, M. Specht, Literature review on co-located collaboration modeling using multimodal learning analytics—can we go the whole nine yards?, IEEE Transactions on Learning Technologies 14 (2021) 367–385.

[3] D. Jonassen, M. Davidson, M. Collins, J. Campbell, B. B. Haag, Constructivism and computer-mediated communication in distance education, American journal of distance education 9 (1995) 7–26.

[4] W. Doise, On the social development of the intellect, in: The Future of Piagetian Theory, Springer US, 1985, pp. 95–121. URL: https://doi.org/10.1007/978-1-4684-4925-9_5. doi:10.1007/978-1-4684-4925-9_5.

[5] J.-W. Strijbos, A. Weinberger, Emerging and scripted roles in computer-supported collaborative learning, Computers in Human Behavior 26 (2010) 491–494.

[6] N. M. M. Dowell, T. M. Nixon, A. C. Graesser, Group communication analysis: A computational linguistics approach for detecting sociocognitive roles in multiparty interactions, Behavior Research Methods 51 (2018) 1007–1041. URL: https://doi.org/10.3758/s13428-018-1102-z. doi:10.3758/s13428-018-1102-z.

[7] L. Menzel, S. Gombert, J. Weidlich, A. Fink, A. Frey, H. Drachsler, Why you should give your students automatic process feedback on their collaboration: Evidence from a randomized experiment, in: O. Viberg, I. Jivet, P. Muñoz-Merino, M. Perifanou, T. Papathoma (Eds.), Responsive and Sustainable Educational Futures, Springer Nature Switzerland, Cham, 2023, pp. 198–212.

[8] D. Di Mitri, J. Schneider, M. Specht, H. Drachsler, From signals to knowledge: A conceptual model for multimodal learning analytics, Journal of Computer Assisted Learning 34 (2018) 338–349.

[9] P. Saint-Andre, Extensible Messaging and Presence Protocol (XMPP): Core, RFC 6120, 2011. URL: https://www.rfc-editor.org/info/rfc6120. doi:10.17487/RFC6120.

[10] H. Bredin, R. Yin, J. M. Coria, G. Gelly, P. Korshunov, M. Lavechin, D. Fustes, H. Titeux, W. Bouaziz, M.-P. Gill, pyannote.audio: neural building blocks for speaker diarization, in: ICASSP 2020, IEEE International Conference on Acoustics, Speech, and Signal Processing, 2020.

[11] H. Bredin, A. Laurent, End-to-end speaker segmentation for overlap-aware resegmentation, in: Proc. Interspeech 2021, 2021.

[12] J. M. Coria, H. Bredin, S. Ghannay, S. Rosset, Overlap-aware low-latency online speaker diarization based on end-to-end local segmentation, in: 2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2021, pp. 1139–1146. doi:10.1109/ASRU51503.2021.9688044.

[13] P. Sobot, Pedalboard, 2021. URL: https://doi.org/10.5281/zenodo.7817838. doi:10.5281/zenodo.7817838.