AME: The Autonomous Multimodal Experiment System

Jiahao Yang[1], *Sotaro Kita*[1]

[1]Department of Psychology, University of Warwick

Jiahao.Yang@warwick.ac.uk, S.Kita@warwick.ac.uk

The modern era of digital communication technologies has expanded the horizons of human interaction by offering various modes of expression, including images, videos, and emojis. These advancements have led to a growing consensus among researchers that the human language faculty transcends the oral modality and is inherently multmodal (Levinson & Holler, 2014; Macuch Silva et al., 2020; Özyürek, 2014, 2021). However, transitioning from a unimodal to a multimodal perspective poses significant challenges in terms of research methodology. Multimodal research, which predominantly deals with audio-visual stimuli and data, has traditionally been conducted in laboratory settings due to the difficulty in collecting, transferring, and storing audio-visual data outside of controlled environments while maintaining satisfactory quality. Consequently, multimodal research has often been limited in terms of sample size and participant demographics. The ongoing COVID-19 pandemic has further complicated matters, as traditional laboratory methods cannot be easily translated into online scenarios.

In response to the pressing need for online multimodal research platforms, we introduce the Autonomous Multimodal Experiment System (AME), a migratable experiment platform that offers an autonomous and innovative solution for conducting interactive multimodal experiments online. The platform is migratable because it can operate on any server and can be accessed via computers with limited processing power and storage capacity; thus it can be adopted by a wide range of users, even those without advanced hardware equipment. The platform's unique built-in workflow facilitates autonomous experimentation by providing automatic field of vision checks through browser-based hand tracking and face tracking (e.g., to make sure the whole upper body is visible). This effectively eliminates the need for experimenter presence during experiment sessions. AME also leverage cutting-edge web technologies such as WebRTC, the VUE framework, and socket.io, as well as browser-based machine learning models to provide distinct functionalities that address the specific requirements of multimodal research. These features are not found in existing online experiment platforms and packages, such as PsychoPy, Gorilla, OTree, and JsPsych (Anwyl-Irvine et al., 2020; Chen et al., 2016; de Leeuw, 2015; Peirce et al., 2019).

Compared to existing platforms, AME has three major advantages. Firstly, AME facilitates rapid collection of high-quality audio-visual data with remarkable efficiency. Its architecture and integrated browser-based machine learning model enable the collection of audio-visual data with sufficient frames per second, an appropriate field of view, and minimal signal noise. Secondly, AME offers unique, innovative functions that make previously impossible experimental designs feasible. It is the first experiment platform to embed browser-based machine learning, video chat, and artificial intelligence language models (e.g., ChatGPT) into behavioral experiments. The ability to assemble these modular components in various combinations unlocks vast potential for experimental design, empowering researchers to tackle new and complex research questions. Lastly, AME provides a seamless experience in constructing, distributing, and managing experiments while prioritizing data privacy. The platform offers a no-code environment, featuring a user-friendly dashboard that allow researchers to design and manage experiments with ease. As a migration-ready platform, AME can also be moved to a private server, which gives researchers full control over their experiments and collected data, ensuring the highest level of data privacy.

To validate the reliability of the AME experiment workflows in ensuring appropriate FPS (frame per second) and FOV (field of vision) of audiovisual data, we conducted an online gesture production experiment using the AME system on a Linux server with a two-core CPU and 4GB RAM. The experiment modified the elicited pantomime paradigm (Ortega & Özyürek, 2020), where participants were asked to depict a highlighted word from a pair of words (e.g., Sandwich and Apple) after a certain time period. AME presented a stimulus word and recorded the participant's gesture depiction by their webcam, resulting in separate video files for each trial. AME ensured the webcam's FOV could capture the participant's upper torso for clear views of the gesture depiction by instructing participants to position themselves from an optimal distance from the camera, using the body-detection function of the AME system.

We collected audiovisual data from 121 adult participants (undergraduate students) recruited in the Department of Psychology at the University of Warwick. Participants produced a total of 9347 depictions, with video durations ranging from 11 to 16 seconds depending on the condition. Using the FFmpeg package in Python, we calculated the frames per second (FPS) for each video in our analysis, excluding videos lacking metadata. This exclusion primarily occurred when participants used the Chrome browser, as Chrome does not set the default duration if the timestamp delta is not consistently constant. Our analysis involved 8084 video files, with an average FPS of 27.47 and a range spanning from 7.5 to 30. Among these videos, 84.3% were recorded beyond 29 frames per second. Figure 1 displays the distribution of FPS in the collected videos. Additionally, we coded the field of view (FOV) for each video, finding that 98% of participant-generated videos (119 out of 121, as exemplified in Figure 2) included the upper torso of the participants. Only one participant failed the FOV check intentionally by blocking the webcam during the position check. Furthermore, one participant's video file was corrupted, likely due to a malfunctioning webcam.The descriptive data confirms the robustness of the AME workflow.

Index Terms: online research, multimodal research, gesture studies

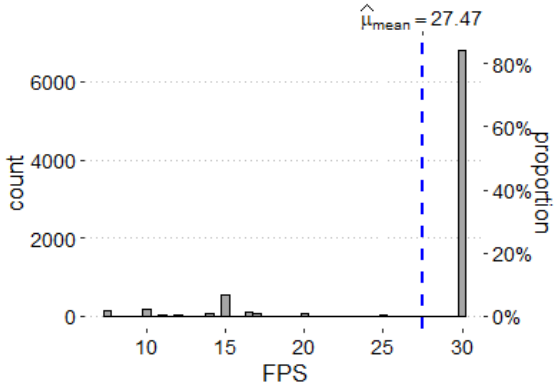

Figure 1: *Distribution of frames per second in collected videos (N=8084). The dash line represents the mean (27.47).*

Figure 2: *Screenshots of videos from the validation stud. Data was collected without experimenter supervision.*

References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. Behavior Research Methods, 52(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. Journal of Behavioral and Experimental Finance, 9, 88–97. https://doi.org/10.1016/j.jbef.2015.12.001

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. Behavior Research Methods, 47(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1651), 20130302. https://doi.org/10.1098/rstb.2013.0302

Macuch Silva, V., Holler, J., Ozyurek, A., & Roberts, S. G. (2020). Multimodality and the origin of a novel communication system in face-to-face interaction. Royal Society Open Science, 7(1), 182056. https://doi.org/10.1098/rsos.182056

Ortega, G., & Özyürek, A. (2020). Systematic mappings between semantic categories and types of iconic representations in the manual modality: A normed database of silent gesture. Behavior Research Methods, 52(1), 51–67. https://doi.org/10.3758/s13428-019-01204-6

Özyürek, A. (2014). Hearing and seeing meaning in speech and gesture: Insights from brain and behaviour. Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1651), 20130296. https://doi.org/10.1098/rstb.2013.0296

Özyürek, A. (2021). Considering the Nature of Multimodal Language from a Crosslinguistic Perspective. Journal of Cognition, 4(1), 42. https://doi.org/10.5334/joc.165

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. Behavior Research Methods, 51(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y