American Journal of
Online and Distance Learning
(AJODL)

Integrating Multimedia Tools to Enrich Interactions in Live Streaming for Language Learning

Eefa Binte Tariq

American Journal of Online and Distance Learning
ISSN 2957-8027 (online)
Vol.7, Issue 1, pp 1 - 13, 2025

www.ajpojournals.org

Integrating Multimedia Tools to Enrich Interactions in Live Streaming for Language Learning

Eefa Binte Tariq

CHI Paper

Article history

Submitted 12.12.2024 Revised Version Received 14.01.2025 Accepted 15.02.2025

Abstract

Purpose: According to the study, multimodal engagement may improve language learning in live-streamed contexts. These include text, speech, video, images, and stickers. The main goal is to examine how new technologies enable more interactive, real-time language instructor-student interactions by overcoming text-only communication in typical online classrooms.

Materials and Methods: One language instructor and nine students participated in a two-week field research project. In four live-streamed lessons, the teacher tested three English teaching methods using text, audio, video, photographs, and stickers. These multimodal channels allow students to engage in class discussions and provide constructive critique actively.

Findings: The study found that multimodal communication engaged students and provided quick feedback. The tools' effectiveness depended on group size, learning context, time of day, and instructors' and students' online identities.

Implications to Theory, Practice and Policy: Online language education may benefit from real-time feedback and involvement via audio, video, text, etc. When creating courses, teachers must consider group size class time, and how students utilize resources.

Keywords: Interactive Experiences; Multimedia; Language Learning; Live Streaming

INTRODUCTION

Educators and students of foreign languages have increasingly leveraged live streaming, a popular participatory social media platform, for language learning. Both private tutors and professional virtual language schools, such as VIPKID, are offering live-streamed sessions to enhance students' learning experiences. During a live stream, viewers can engage in real-time textual communication with the streamer through a chat channel, often contributing hearts and comments. Initiating a live stream on most social media platforms is user-friendly, typically requiring just a few clicks, and viewers can join or leave the stream at any time. The live audience can range from a few individuals to tens of thousands, providing an interactive and global language-learning experience (Richard, 2018).

The key advantage of live streaming in distance language education is the two-way interaction between instructors and students. However, most live-streaming platforms currently limit user interactivity to text comments and basic emojis. While these forms of communication are quick and easy, they do not support essential language learning components like speaking skills. Given that language is inherently multimodal, with multiple ways to express oneself, learning is believed to benefit from multimodal stimuli. In the field of Computer-Assisted Language Learning (CALL), considerable research has focused on integrating multimedia resources—such as audio, images, and videos into language learning. Studies in CALL have demonstrated that multimedia content provides richer cues, aiding in language acquisition, particularly in areas like pronunciation and vocabulary development (Bakhshi et al., 2016).

This study explores the use of multimedia technologies in real-time live streaming as a tool for language acquisition in small-group settings. To enhance live-streamed language classes, we incorporated various comment formats, including text, audio, video, images, and stickers. Although these modalities have been widely used in other social media platforms like MMS and MIM, their potential applications in live-streamed language learning have not been extensively studied. The following research questions guided our study:

- How do multimedia tools in live streaming affect interaction between students and teachers?
- How engaged are students in live-streamed classes?
- How do these tools influence teaching methods and student learning outcomes?

To address these questions, we conducted a two-week longitudinal study in which nine participants attended four English lessons taught by a streamer using a multimodal live-streaming system. We employed a diary research approach, administered post-session questionnaires, and concluded with interviews to examine how both the streamer and the viewers utilized multimodal comments during language learning. This study explores several aspects of multimedia-enhanced live streaming for language learning, including stream management, factors influencing the streaming experience, interaction dynamics between the streamer and viewers, and insights into how multimedia tools facilitate language acquisition.

Problem Statement

The lack of platforms and research has prevented language education from using live streaming. Most of what is known about language learning is focused on text-based interactions, not multimodality. Asynchronous Computer-Assisted Language Learning (CALL) has proven that multimedia tools like audio for pronunciation and video for visual cues are beneficial, but these results have not been applied to interactive, real-time environments. Research on multimodal live streaming, including Twitch, emphasizes entertainment-driven engagement rather than pedagogical needs, which conflicts with educational aims and technical capabilities. This difference prevents students and instructors from using tools that may substitute face-to-face encounters, when body language, voice, and pictures promote comprehension and memorization.

Previous research identified three major drawbacks. First, live streaming services concentrate on text/emoji communication and overlook audio/video aspects that are essential for speaking and listening skills, thus there is no real-time feedback on pronunciation and nonverbal cue exchange. Second, there is little study on the possibility of dynamic, interactive education when multiple modalities are merged in real-time, which is essential for quick feedback, spontaneous discourse, and collaborative problem-solving. This contrasts with CALL's well-documented usage of asynchronous multimedia like taped movies. Third, contemporary multimodal technologies in non-academic contexts (such as push-to-talk in games) lack pedagogical frameworks for language training, rendering them unsuitable for structured skill development. Due to the lack of research on small-group dynamics in live streaming environments, which is ideal for nuanced language practice, and the difficulty of providing personalized feedback to large audiences, educators lack evidence-based methods to balance class size and interaction quality.

To fill these gaps, this longitudinal, in-the-wild study examined the real-time effects of text, voice, video, images, and stickers on language acquisition. HCI/EdTech researchers increase knowledge, while students benefit from immersive, multimodal practice that simulates real-world communication. Teachers gain from adaptive teaching practices that limit one-way instruction. Empirical findings help platform designers enhance comment management and moderation. Integrating CALL concepts with live streaming innovation, this innovative project creates practical frameworks to improve digital language learning. Creating more engaging and effective learning experiences is the objective.

LITERATURE REVIEW

This section reviews significant previous research on key topics, including multimodality, language acquisition, computer-mediated communication (CMC), and live broadcasting.

Multimodality in Language Learning

Prior research has extensively explored multimedia learning tools that enhance communication. For instance, Hayashi et al. proposed a user-friendly multimodal interaction platform for facilitating computer-assisted group study, while Yoon’s multimodal annotation system enables students to collaborate remotely through a combination of speech, writing, and gestures to discuss ideas.

In linguistics, the use of multiple modalities for second language acquisition remains a prominent area of interest. Multimodal components have been integral to foreign language teaching methods for quite some time. Children's books, for example, often demonstrate a strong synergy between text and illustrations. Additionally, Gilakjani et al. identified key principles of multimodal learning, highlighting its positive impact on language acquisition.

More recently, research has focused on integrating multimedia elements into video-based language learning. Zhu et al.'s video-augmented dictionary, which uses internet videos to enhance vocabulary learning, is one such innovation (Anon, 2014). Technological advancements have also made language learners more pragmatic. These programs allow learners to search videos by context and expression, and systems like Seiyuu-Seiyuu use voice-driven features to help users practice pronunciation from any video. Past studies have shown that multimedia content, when paired with text, provides more nuanced cues for learners, improving their understanding and retention of language material.

Language-Learning and Computer-Mediated Communication

Since the introduction of Computer-Assisted Language Learning (CALL) in the 1960s, digital technology has become a key part of foreign language education. The growth of the internet and the rise of distance learning have allowed language learners to access vast amounts of information remotely through a variety of online platforms. In particular, synchronous forms of Computer-Mediated Communication (CMC) like online video and audio conferences, as well as live broadcast lectures, have proven especially beneficial for language acquisition (Austin, 2024).

Audio and video conferencing are well-known tools for practicing speaking and listening skills. Platforms such as Skype and Google Hangouts enable language instructors and students to engage in real-time, two-way conversations. In addition, live-broadcast lectures allow remote participants to interact by submitting written comments, voting in polls, or using a "raise hand" feature to ask an audio question. Systems like TELEP support simultaneous participation from both local and remote audiences, facilitating live exchanges much like traditional classroom settings. In these scenarios, participants often take turns asking questions, similar to in-person classes.

While these synchronous CMC technologies bear similarities to live streaming, live streaming offers a unique space for language learning by fostering immediate audience interaction and allowing the public to influence the content of the stream in real-time.

Multimodality in Live Streaming

Live streaming engagement relies heavily on interaction and social dynamics. Researchers have explored multimodal live-streaming communication to enhance viewer-streamer engagement. Hamilton et al. developed a highly visible and instantly engaging live-streaming system using push-to-talk (PTT) audio. Lessel and colleagues introduced a novel communication approach for Hearthstone, allowing the audience to influence the streamer through additional communication modes. Polls on Twitch during live streams help streamers make critical in-game decisions. Since platform-integrated text chat systems cannot perform this function, streamers often use third-party applications for polling viewers. Outpost Games designed Hero.tv to improve streamer-audience interaction for their game SOS, allowing supporters to drop power-ups into the game environment when a broadcaster fires a flare. Platforms like Facebook Live, YouNow, and Live. I offer guest

broadcasting features that enable real-time viewer interaction with streamers. Prior research indicates that multimodal interaction in live streaming is feasible and can improve the viewing experience. This study examines how multimodal communication is used in live-streamed language courses (Kulkarni et al., 2015).

Multimodal technologies may aid language acquisition, but research has seldom linked them to specific objectives like fluency, comprehension, or vocabulary retention. For instance, live streaming services cannot offer feedback like this, which prevents fluency-building real-time pronunciation practice. Text-centric streams fail to explain phonetics and idiomatic idioms like video interactions, which involve lip motions and gestures. Sharing images helps pupils recall phrases, but there is no systematic way to use live broadcasts to generate descriptive activities using photos. Despite being understudied, stickers indirectly improve fluency by encouraging hesitant learners to participate, reduce fear, and increase engagement.

Asynchronous resources like prepared films are fantastic for self-paced learning, but they can't teach practical skills like turn-taking, which are best learned via natural conversation. Twitch doesn't consider the educational possibilities of polls and push-to-talk for vocabulary drills, while video conferencing and other synchronous platforms give interactive fluency but can't scale. Live streaming's scalability makes things worse: small-group dynamics, ideal for nuanced practice, lack study, while large audiences make tailored feedback difficult.

This research examines how synchronous multimodal live-streaming systems try to address these issues:

- Auditory and visual pronunciation feedback.
- Video and images may enhance understanding.

Picture-based vocabulary reinforcement in context.
This research coordinates resources and outcomes to encourage technology-based language learning.

Method

In our study on using live streaming as a medium for language acquisition, we examined the integration of multimedia resources over two weeks with a single broadcaster and nine viewers. We chose a smaller audience to facilitate more personalized, one-on-one interactions with each participant and to understand how multimodal technologies could enhance relationships. Viewers engaged in the language lessons through various commenting options while the streamer utilized a live-streaming program. The goal was to observe how the streamer employed multimedia interactions to improve her lessons and how the viewers used these tools for enhanced learning.

To minimize disruption to the streamer's usual routine, we gave her considerable flexibility in determining the number of live streams, their length, and content. The only request was that she maintain a schedule of two to seven live broadcasts each week to ensure sufficient data collection. Throughout four 30-minute live broadcasts, the streamer experimented with three different instructional approaches. During these sessions, one researcher joined the stream to record the content for later review, providing technical assistance when necessary but otherwise remaining unobtrusive to the lessons.

Table 1: Viewer Participant Details

Viewer	Age Range	Gender	Occupation	Native Language	Self-Reported English Proficiency	Was Streamer's Existing Student
P1	26-30	Male	Customer Service	Spanish	Advanced	No
P2	41-45	Male	Personal Trainer	Punjabi	Intermediate	No
P3	36-40	Male	System Analyst	Portuguese	Advanced	Yes
P4	31-35	Male	Junior High School Teacher	Chinese	Intermediate	No
P5	21-25	Female	Digital Marketeer	Portuguese	Intermediate	Yes
P6	31-35	Male	Higher Education English Teacher	Spanish	Intermediate	Yes
P7	26-30	Female	Kindergarten and Elementary School English Teacher	Chinese	Intermediate	Yes
P8	41-45	Female	Nanny	English	Fluent	No
P9	26-30	Female	Software Engineer	Korean	Intermediate	No

Participants

To recruit a suitable streamer for this study, a search was conducted across popular video-sharing and live-streaming platforms such as Facebook, YouTube, and Periscope. The selected streamer had extensive experience providing live English language lessons through a variety of formats, including live broadcasts, group video chats, and pre-recorded sessions. By the time of the research, she had been streaming for nearly a year, primarily using platforms like Facebook Live and Instagram Live for her classes. During the initial setup interview (see Section 3.3), the streamer shared insights into how she engaged with her audience in real-time, adapting her lessons to meet the needs and preferences of her viewers.

To recruit participants for the study, current students of the streamer were contacted, and advertisements were placed on social media and other platforms to attract new viewers. Participants were required to have previous experience watching live-streamed English language lessons, as the study focused on intermediate to advanced English speakers. Only one beginner-level speaker was recruited, but they later chose not to participate, potentially due to the linguistic demands of following a live-streamed lesson and engaging with other students. While most participants had a strong command of the English language, one participant (P8) was particularly motivated to refine her British accent, which she practiced during her free time by watching live-streamed lessons (Church & de Oliveira, 2013).

System

A smartphone app was initially developed to facilitate multimodal commenting during live-streamed sessions. However, a preliminary trial with participants revealed that the app lacked the reliability needed for seamless live broadcasting. Consequently, Facebook Live and Facebook Messenger were used for the official study. Facebook Live was chosen for its stable streaming capabilities, while Messenger was selected for its ability to support a wide range of commenting

options. These platforms were preferred over others due to their user-friendly interfaces and the specific features they offered, which aligned with the goals of the research.

To initiate a live broadcast, the streamer accessed her Facebook page, selected the Live button, chose her audience, added a description, and then started the live video. Participants could join the stream by clicking on the Facebook notification. The broadcasts were private and accessible only to the study's participants (see Figure 1, left). While watching the stream on their desktop computers, participants simultaneously engaged in a Facebook Messenger group chat using their smartphones. Messenger provided various commenting options, including text, voice, video, images, stickers, and Facebook "like" reactions (see Figure 1, right). To send an audio comment, participants held down the microphone button while speaking. They could also share images or videos by using the app's camera function or uploading files from their device's local storage. A scrollable panel of stickers and emoticons allowed participants to express themselves further. The streamer and participants could interact with all comments, which were available for review or playback at any time. Audio and video comments from the streamer were broadcast to all viewers in real-time via the live-streaming device. Participants were instructed to communicate exclusively through Messenger, rather than using the Facebook Live comment section, to streamline interactions (Corbin & Strauss, 2018).

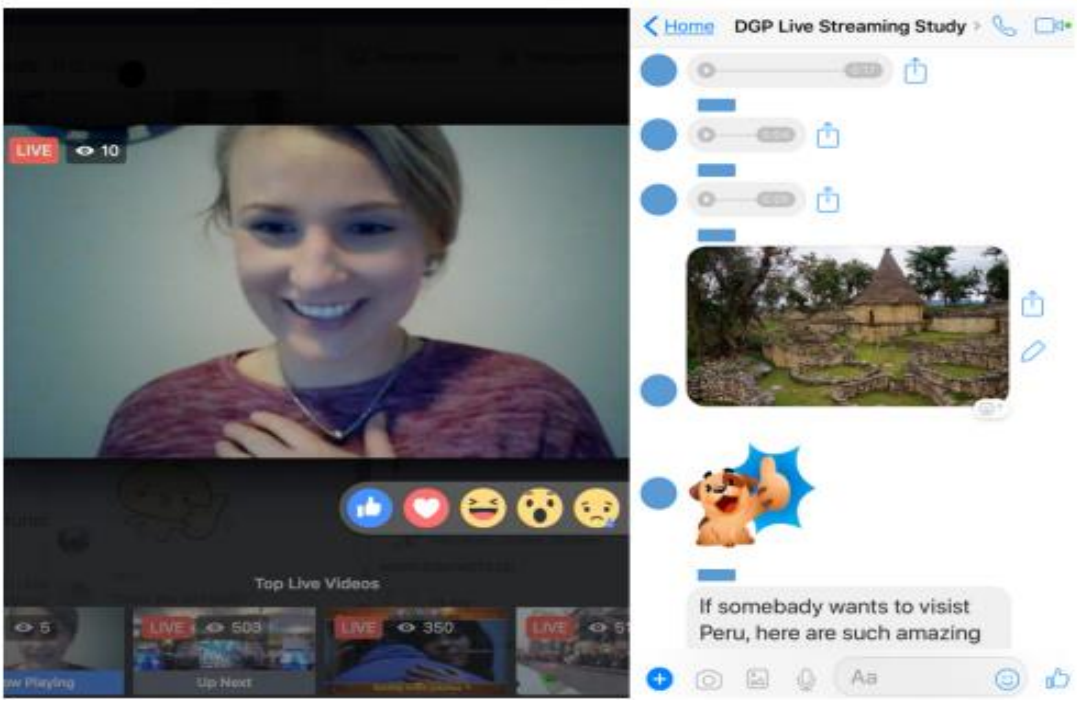

Figure 1: The Language Lessons Were Live-Streamed on Facebook Live (Left), while The Multimodal Comments Were Sent on Facebook Messenger (Right). The Right Figure Shows Three Audio Comments, Followed by an Image, A Sticker, and A Text Comment. Participants' Name

Multimedia technologies in live streaming for language learning are discussed, along with internet speed and streaming system stability. It describes research challenges but does not specify internet speeds or video resolutions:

1. Technical Challenges: Since their original mobile app for the study was problematic, the researchers used Facebook Live and Messenger. The streaming experience's stability determined our choice. However, background noise distorted audio when viewers recorded audio comments with the live stream.
2. Overcoming Technical Issues: Participants were taught to manually silence the live broadcast or utilize headphones. Consumers need technological adaptability for real-time communication.
3. Although technological reliability is crucial, the paper discusses how these concerns affect communication and participation rather than technical characteristics like internet speed or video quality.

The study emphasizes the need to successfully manage technical difficulties like internet connectivity and audible sound to fully engage with multimodal live streaming settings for language learning.

Live Stream Sessions

The formal analysis was carried out by the streamer using three distinct educational approaches. Two of the four live sessions focused on pronunciation drills, where the streamer addressed challenging English words and phrases with the assistance of viewers who were encouraged to record their pronunciations and receive feedback. In another live broadcast, the streamer facilitated a conversational activity, initiating dialogue with viewers by asking questions like, "What did you have for breakfast?" The conversation was then passed between viewers, with each responding to the streamer or the previous viewer via audio or video comments. Between conversation topics, the streamer introduced idioms, expanded the vocabulary, and offered pronunciation corrections.

In the final live session, participants were asked by the streamer to describe images. Viewers were invited to submit images of their hometowns or countries and provide detailed descriptions of the scenes, which contributed to an interactive learning experience (see Table 2).

After each live broadcast, participants completed a questionnaire to reflect on their use of multimedia technologies during the session. These questionnaires acted as "diaries," allowing participants to share their thoughts and provide insights into their experiences. Two versions of the surveys were created—one short and one long—each tailored to gather specific feedback from both viewers and the streamer. The shorter version, used after the first and third live sessions, focused on questions related to communication modes, such as how comfortable participants felt using various forms of comments, and their thoughts on particularly significant interactions or

emotions. This brief version was kept concise to maintain participant focus and prevent fatigue (Culbertson et al., 2017).

The longer questionnaire, administered after the second and fourth live sessions, retained the questions from the shorter version but added queries related to the learning experience, engagement, and communication with both the streamer and other viewers. These extended surveys were used at key points in the research to assess more abstract factors such as participant engagement and overall learning, which required data collected over a series of broadcasts.

Surveys tailored for the streamer followed a similar format but were adjusted to reflect her experience throughout the project. In cases where participants missed a live session due to

scheduling conflicts, they were encouraged to view the archived version of the stream to stay up-to-date with the lessons. Table 2 provides an overview of participant numbers for each live session.

The study analyses interaction, participation, and experiences during live-streaming language classes, but it doesn't evaluate learning outcomes like fluency, vocabulary, or pronunciation.

- Interactives Matter More Than Outcomes: The research examined instructor-student interactions and multimedia tools. There was no systematic evaluation of participants' pre- and post-session language proficiency.
- Improvements: A pre-and post-test for key language skills would improve dependability. Proper pronunciation, vocabulary, and speaking fluency are included. This may help measure the streaming sessions' interactive methods' learning benefits.
- Step Three: Future research may benefit from adding pronunciation software or vocabulary exams before and after the session to better assess learning outcomes. Results would indicate how effectively multimodal interactions aid learning.
- Finally, although the study illuminates the social aspects of language acquisition via live streaming, it may benefit from assessing learning outcomes to support its results and demonstrate its efficacy.

Table 2: Language Lesson Livestreaming Details. Lessons are Labeled L1-L4

Viewer	Age Range	Gender	Occupation	Native Language	Self-Reported English Proficiency	Was Streamer's Existing Student
P1	26-30	Male	Customer Service	Spanish	Advanced	No
P2	41-45	Male	Personal Trainer	Punjabi	Intermediate	No
P3	36-40	Male	System Analyst	Portuguese	Advanced	Yes
P4	31-35	Male	Junior High School Teacher	Chinese	Intermediate	No
P5	21-25	Female	Digital Marketeer	Portuguese	Intermediate	Yes
P6	31-35	Male	Higher Education English Teacher	Spanish	Intermediate	Yes
P7	26-30	Female	Kindergarten and Elementary School English Teacher	Chinese	Intermediate	Yes
P8	41-45	Female	Nanny	English	Fluent	No
P9	26-30	Female	Software Engineer	Korean	Intermediate	No

	Teaching Strategy	Questionnaire Completed	Size of Real-Time Audience
L1	Pronunciation	Short	6
L2	Pronunciation	Long	9
L3	Conversation	Short	6
L4	Picture Description	Long	5

Final Interviews

Following the completion of the two-week study, all participants were interviewed via semi-structured video chats to gain insights into their use of multimodal comments during real-time language classes. Each interview lasted approximately thirty minutes. Questions were tailored to align with participants' responses from the questionnaires. Viewers were asked to evaluate their engagement, motivation, and involvement in learning, as well as how multimodal live streaming compared to traditional live streaming. They were also asked to describe their use of each commenting modality, their interactions with the streamer and other viewers, and how multimodal commenting impacted their learning experiences. Additionally, participants were prompted to reflect on the effectiveness of the three teaching strategies employed during the study. Similar questions were posed to the streamer, along with inquiries about the tools used to track participants' progress and facilitate English language instruction. To better understand emotional responses, participants were encouraged to share personal experiences and the context behind specific feelings.

After the trial, participants were entered into a CAD 100 gift card drawing. The streamer was compensated with 96 CAD, while each viewer received 30 CAD for their time.

Data Analysis

All interview data were analyzed using an open-coding approach. Questionnaire responses were not open-coded, as planned, but participants were asked to expand on their questionnaire answers during the final interviews. Researchers identified key phrases, sentences, and paragraphs within the interview transcripts and labeled the data accordingly. After independently coding 20% of the data, two researchers met to reach a consensus on the codes. Following this, a single researcher completed coding the remaining data. Finally, the team aggregated the transcripts and developed themes and subthemes using affinity diagrams.

The study's consensus coding technique required two researchers to code the first 20% of interview material independently to produce initial codes. One researcher's subsequent classification of the remaining data raises bias concerns. Peer validation improves research objectivity and openness. Another researcher examined the coding and interpretations to ensure no biases affected the findings. A more comprehensive research may benefit from triangulation, which entails collecting data from many sources and cross-referencing the findings. Researchers should reflect on their prejudices and experiences to put their opinions in context. Collaboration coding, in which a group of academics categorizes data individually before comparing and contrasting their conclusions, may enhance future research by providing several perspectives. These methods would increase the

study's rigor and trustworthiness. Fixing these flaws makes the research findings more trustworthy and helps influence MLE best practices.

Table 3: Frequency of Multimodal Comments in Live Courses

	Text	Audio	Video	Image	Stickers	FB Like	Total
L1	89	26	4	0	5	4	128
L2	56	55	2	0	21	7	141
L3	50	47	1	9	12	6	125
L4	44	40	0	5	9	0	98

Findings

We identified five major themes in our data: the use of multimedia technologies in live-streamed language classes, streamer-viewer interactions, factors influencing multimodality, motivation, and engagement, and lesson design and management.

Multimedia Technologies in Live-Streamed Classes

As shown in Table 3, text comments were the most frequently used modality during live broadcast sessions. Viewers utilized text for various purposes, such as greeting the streamer, asking lesson-related questions, providing feedback, and staying engaged with the lesson content. The text was particularly valuable during pronunciation lessons, where the streamer typed out words to help maintain focus (P2). Text comments also functioned as a means of self-motivation, with participants using them to monitor vocabulary, especially when new or complex terms were introduced. For example, when one viewer typed an unfamiliar word, another viewer noted that they were able to learn its meaning and spelling by seeing it in text (P3).

Audio comments were the second most popular modality, especially during pronunciation classes, where students recorded and submitted audio clips for feedback from the streamer. This created an iterative process where viewers would practice, receive corrections, and attempt again. However, due to time constraints, not all voice comments could be addressed. Audio was also widely used in conversation practice, allowing participants to engage in English dialogue with each other (P9) and work on connected speech (P5). Audio comments helped participants express emotions, clarify questions, and share descriptions of images in the chat group (P6).

Streamer-Viewer Interactions

Participants found that video comments offered a unique advantage, particularly in pronunciation practice. Videos allowed the streamer to observe the viewers' mouth movements, tongue placement, and overall pronunciation technique, which greatly enhanced feedback (P5, P7). However, despite its benefits, the video was used less frequently due to technical challenges and the additional time required for recording and uploading (P2). Occasionally, videos were shared

to provide cultural context, such as weather updates or local scenery, which helped foster a sense of community and motivation among viewers (P6).

Images also played a role in encouraging linguistic development and sparking conversations. Describing images helped participants practice their speaking skills, even though the images themselves did not directly contribute to language acquisition (P5). Images provided an opportunity to introduce new topics and expand cultural awareness, as participants shared pictures from their own countries (P7). However, preparation was often necessary, with participants needing to search for suitable images online during the live broadcast (P2, P7).

Factors Influencing Multimodality

Audio and stickers were consistently identified as the most effective modalities for communication and learning. Given the focus on speaking in every live session, audio provided the best means for viewers to demonstrate their progress, while stickers were easy to use, enjoyable, and helped build rapport among participants. The use of images and videos, while beneficial, was less frequent due to the additional effort required and the technical issues encountered. One streamer noted that although audio and video were useful for feedback, they were also exhausting to manage (P6).

Motivation and Engagement

Multimedia technologies allowed viewers to assess their areas of improvement in language learning. For instance, when discussing image descriptions, one participant reflected on their lack of vocabulary but recognized the experience as a valuable opportunity to identify areas needing more practice (P5). The combination of modalities allowed for deeper engagement, encouraging participants to actively participate in lessons through diverse forms of communication.

Lesson Design and Management

The use of multimodal live-streaming interactions highlighted innovative approaches to language learning, allowing the streamer to experiment with various teaching strategies while engaging with viewers in real time. These interactions not only facilitated feedback but also created a dynamic learning environment where participants could identify personal challenges, such as vocabulary gaps, and work toward improvement.

Discussion

The research highlighted that multimodality significantly enhanced learning by providing fast feedback and improving interactions during live broadcast language lessons. Learners recognized that smaller audiences benefited the most from multimodal instruction via live video, as the streamer could engage more personally with each participant. In larger settings, this individual attention becomes difficult, mirroring the challenge of traditional classrooms where the instructor's attention is divided among many students.

To address this, moderators could play a crucial role in helping manage live-streamed classes, especially with larger audiences. Like Twitch broadcasting communities, moderators could welcome viewers, answer questions, and facilitate engagement. In an educational context, moderators could act as teaching assistants (TAs), helping the streamer respond to questions and encouraging active participation, which becomes harder to maintain with a larger group.

Another suggested strategy is to split the audience into smaller interaction groups. Prior studies have shown that small-group video chats can enhance engagement and learning in large, global classrooms. In live streams, moderators could create smaller comment groups, where participants can discuss and ask questions more easily. The streamer could also alternate between focusing on smaller groups and addressing the whole audience for broader discussions. However, this requires effective comment management and group organization, which remains a challenge to be explored.

While the study focused on language learning, which is a common topic in live streaming, the findings may not be directly applicable to other forms of content like mathematics, coding, or yoga instruction. Additionally, the results may not generalize to larger-scale online learning platforms or tutoring sessions, as this research focused on a relatively small and diverse sample group. Factors such as demographic differences, technological proficiency, and cultural backgrounds might have influenced the participants' use of multimodal technologies.

Research shows that multimodal engagement during live language broadcasting gets harder to scale as the audience rises. Due to relationship management issues with larger groups, viewers may feel ignored and uninvolved. When a broadcast has a lot of comments, it might drain the streamer's time and energy, lowering their reaction and engagement. To resolve these difficulties, specific solutions like AI-driven moderation systems may be beneficial. These technologies may automatically sort and filter viewer comments to recognize them. This would allow broadcasters to focus on key discussions. Additionally, streamers may be able to detect uninterested viewers using AI to measure their engagement levels in real-time. Breakout rooms for smaller group discussions are another option. Moderators might keep the session flowing by facilitating smaller groups. These practical methods may greatly increase multimodal interaction scalability, making education more engaging and adaptable for all learners in a variety of circumstances.

Comparing the findings to digital language learning research may help increase the multimodal involvement in live language streaming discussions. Fuller interactions employing multimodal technology increase learning, supporting previous studies showing that multimedia aids language acquisition. However, Zhu et al. found that standard video-based resources may disengage students owing to the lack of interaction. Real-time interactions in live streaming systems have several advantages, as this comparison illustrates. This study emphasizes active engagement, like Yoon's multimodal annotation systems, which allow students to collaborate across several channels. This study could address the fact that different digital learning platforms have different outcomes to contribute to a larger conversation about how to optimize multimodal approaches for different learning environments based on Church and de Oliveira's previous research that compares communication modes. These references would contextualize the study's results and highlight its unique contributions to digital language learning topics.

Conclusion

This study used an empirical approach to examine how live-streaming multimedia affects student-teacher interactions, engagement, and communication in language learning. It is the first longitudinal in-the-wild research to explore the use of multimodality in live-streamed language lessons. While questions remain about scaling multimodal engagement for larger audiences, the research shows that multimodal interactions foster more natural, personal engagement, and

promote active learning. Results of multimedia technology in live-streamed language classrooms show potential for boosting language acquisition. Text comments are important for communication, but their ubiquity makes it hard to provide meaningful feedback, particularly on pronunciation and public speaking. This suggests that although the text is needed for vocabulary aid and explanation, audio, and video should be utilized more to engage. To encourage a more holistic approach to language learning, instructors should actively promote these richer modalities. For pronunciation practice, audio comments, which allowed for iterative feedback loops between the streamer and learners, worked well. This interactive approach improves learning and creates a safe place for mistakes. Thus, language instructors should prioritize developing mechanisms for easy audio submission and playback to improve this kind of engagement. The varied degrees of participation also demonstrate the necessity to accommodate active and passive learning roles. Understanding these obligations helps instructors create engaging strategies for all students. To make viewers feel less alone, scheduled breakout sessions may foster peer relationships. Finally, the study emphasizes the necessity to balance educational fun with active involvement. Pleasure significantly affects language learning motivation and persistence. Future language learning experiences should stress establishing a comfortable location for students to study with appealing resources. Finally, our findings enhance online language learning training and demonstrate how innovative ways might boost student interest and performance.

REFERENCES

Anon. (2014, October 4). *How To Make Learning Relevant To Your Students (And Why It's Crucial To Their Success)*. Open Colleges. https://www.opencolleges.edu.au/blogs/articles/how-to-make-learning-relevant-to-your-students-and-why-it-s-crucial-to-their-success

Austin. (2024). *Computer-Mediated Communication | Foreign Language Teaching Methods: Writing*. Utexas.edu. https://coerll.utexas.edu/methods/modules/writing/01/cmc.php

Bakhshi, S., Shamma, D. A., Kennedy, L., Song, Y., de Juan, P., & Kaye, J. 'Jofish'. (2016). Fast, Cheap, and Good. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. https://doi.org/10.1145/2858036.2858532

Church, K., & de Oliveira, R. (2013). What's up with WhatsApp? *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '13*. https://doi.org/10.1145/2493190.2493225

Corbin, J., & Strauss, A. (2018). *Basics of Qualitative Research (3rd ed.): Techniques and Procedures for Developing Grounded Theory - SAGE Research Methods*. Sagepub.com. https://methods.sagepub.com/book/basics-of-qualitative-research

Culbertson, G., Shen, S., Jung, M., & Andersen, E. (2017). *Facilitating the Development of Pragmatic Competence through a Voice-driven Video Learning Interface*. https://doi.org/10.1145/3025453.3025805

Kulkarni, C., Cambre, J., Kotturi, Y., Bernstein, M. S., & Klemmer, S. R. (2015). *Talkabout*. https://doi.org/10.1145/2675133.2675166

Richard. (2018). *Allwright, R. (1984). The importance of interaction in classroom language learning. Applied Linguistics, 5, 156-171. - References - Scientific Research Publishing*. Scirp.org. https://www.scirp.org/reference/referencespapers?referenceid=1043054