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

Di (Laura) Chen
University of Toronto
Toronto, ON, Canada
chendi@dgp.toronto.edu

Dustin Freeman
Escape Character Inc.
Toronto, ON, Canada
dustin@escape-character.com

Ravin Balakrishnan
University of Toronto
Toronto, ON, Canada
ravin@dgp.toronto.edu

ABSTRACT

Online language lessons have adopted live broadcasted videos to provide more real-time interactive experiences between language teachers and learners. However, learner interactions are primarily limited to the built-in text chat in the live stream. Using text alone, learners cannot get feedback on important aspects of a language, such as speaking skills, that are afforded only by offering richer types of interactions. We present results from a 2-week in-the-wild study, in which we investigate the use of text, audio, video, image, and stickers as interaction tools for language teachers and learners in live streaming. Our language teacher explored three different teaching strategies over four live streamed English lessons, while nine students watched and interacted using multimodal tools. The findings reveal that multimodal communication yields instant feedback and increased engagement, but its use is dependent on factors such as group size, surroundings, time, and online identity.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; *Collaborative and social computing*.

KEYWORDS

Language Learning; Live Streaming; Multimodal Communication; Multimedia; Interactive Experiences

ACM Reference Format:
Di (Laura) Chen, Dustin Freeman, and Ravin Balakrishnan. 2019. Integrating Multimedia Tools to Enrich Interactions in Live Streaming for Language Learning. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK*. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3290605.3300668

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CHI 2019, May 4–9, 2019, Glasgow, Scotland UK
© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300668

1 INTRODUCTION

The rise of live streaming as a popular form of participatory social media has attracted language educators and learners to take advantage of this interactive platform. Professional virtual language education services such as VIPKID [45] and an increasing number of independent educators like [35] are offering live streams for language learning. A live stream typically involves a streamer broadcasting in real-time to viewers who can send text comments and hearts through a synchronized chat channel. On most social networks, anyone can start broadcasting a live stream with just a few clicks of buttons, and viewers can drop-in to watch the stream or leave the stream at any time. The real-time audience can range from a few people to tens of thousands of viewers. For language learning, this convenient technology allows ordinary people to connect with others from around the world and engage in real-time shared learning experiences.

One of the key aspects of live streaming that makes it an appealing platform for online language learning is the interactivity between the teacher and the student viewers. However, viewer interactions in most existing live streaming services are confined to text comments and simple emojis. While text and emojis serve as quick and simple ways of communication, they are limited in their capacity to facilitate the learning of fundamental language skills, such as speaking. Language is multimodal by nature, as speech is invariably accompanied by multiple channels of expression [44]. Multimodal cues are argued to be important for language acquisition [13, 30]. The use of multimedia materials such as audios, images, and videos has also been extensively studied in Computer-Assisted Language Learning (CALL). Research in CALL have demonstrated that different aspects of language learning, such as pronunciation and vocabulary building, have benefited from the richer cues afforded by multimedia content [47, 51].

In this work, we study the usage of several multimedia tools in live streaming for language learning with an intimate class size. In addition to the basic text comments, we incorporated audio, video, image, and stickers comments into the

live streamed language lessons. Although these modalities are already well-established in other social media such as Multimedia Messaging Service (MMS) and Mobile Instant Messaging (MIM) [5], their usages have not been explored in the context of language learning in live streaming.

We focused on the following research questions — how do multimedia tools in live streaming for language learning affect:

(1) teacher-student interactions and peer interactions?
(2) student engagement in the live streamed lessons?
(3) the teacher's teaching strategies and students' learning experiences?

To investigate these questions, we conducted a 2-week longitudinal in-the-wild study during which we recruited a streamer to teach four English lessons on a multimodal live streaming system, and nine viewer participants to watch the live streams. We observed how the streamer and viewers used multimodal comments for language learning by employing a diary study approach [10], asking participants to fill out a questionnaire after each live stream session, and conducting a final interview. We present findings on the usage of multimedia tools in live streaming for language learning, the manageability of the live stream, factors affecting the streaming experience, communication and interactions between the streamer and viewers, and insights on multimedia-enriched live streaming for language learning.

2 RELATED WORK

In this section, we discuss prior relevant works on multimodality, language learning, Computer-Mediated Communication (CMC), and live streaming.

Multimodality in Language Learning

Prior research has widely explored the use of multimedia materials in learning to afford more enriched and clear communication. Hayashi et al. [18] proposed an accessible multimodal interaction platform for a computer-supported collaborative learning system. Yoon [52] developed a multimodal annotation system that allows students to exchange ideas remotely using combinations of voice, text, and pointing gestures.

Multimodality in *language* learning has also been a topic of great interest in linguistics literature. Instructional practices used in foreign language learning and teaching have always included a multimodal dimension [9]. In children's books, we find that often, text and picture complement each other [9]. Gilakjani et al. [12] have also identified important principles of multimodal learning and their positive effects on language acquisition. Recent research has explored the use of multimedia material in video-based language learning. For example, Zhu et al. [55] presented a video-augmented dictionary that incorporates existing online videos for vocabulary learning. To help language learners develop pragmatic competence, video learning tools have been enhanced: the voice-driven Seiyuu-Seiyuu system [7] enables learners to practice saying phrases from any video, and Exprgram [24] supports context- and expression-based browsing using learner-sourced video annotations. Prior works have shown that a combination of multimedia content in language learning provides richer cues than text alone, and can promote more effective interpretation and memorization of the learning materials.

Language Learning and Computer-Mediated Communication

Digital technologies have become an integral part of foreign language learning since the introduction of CALL in the 1960s. The widespread use of the Internet has opened up opportunities for distance learning, enabling language learners to acquire knowledge remotely through various online resources. In particular, synchronous CMC, such as audio or video conferencing, network broadcasted talks, and live streaming, offer real-time interactions between language educators and learners, which can contribute significantly to the learning experience.

Audio or video conferencing are well-known approaches to practicing speaking and listening skills [17, 29]. Through services such as Skype or Google Hangouts, language learners and teachers could establish real-time, bi-directional audio and visual connections. In network broadcasted talks, the speaker could broadcast live audio, video, and slides over a network, while remote viewers could send written comments, vote on a poll, and "raise their hand" to ask an audio question by clicking on a button [21]. Some systems, such as TELEP, support the co-presence of both local and remote audiences [22]. Network broadcasted talks usually require the audience to take turns when interacting with the speaker, mimicking the style of a classroom lecture. Live streaming shares similarities with these existing synchronous CMC methods. However, live streaming's affordances of simultaneous audience participation and its ability for the audience to influence the content of the stream make it a unique platform to explore for language learning.

Multimodality in Live Streaming

Two crucial properties that contribute to engagement in live streaming are interaction and sociality [14]. To establish better viewer-streamer interactions, researchers have explored various forms of multimodal communication in live streaming. Hamilton et al. [16] incorporated push-to-talk (PTT) audio into their live streaming system, which proved to be an engaging modality due to its instant and high-profile nature. Lessel et al. [28] built several new communication channels for the live streamed card game Hearthstone. They

found that additional communication modalities are valuable to the audience due to the influence they can exert on the streamer. On Twitch, polls are often conducted during live streams to help the streamer make critical decisions in a game [15]. Since platform-integrated text chat systems cannot properly support polling, many streamers choose to use third-party tools to determine the viewers' preferences [15, 28]. Outpost Games built the Hero.tv platform to enable more in-depth streamer-audience interaction for their game SOS; in SOS, when a streamer fires a flare, audience fans can drop helpful power-ups into the game world at that location [38]. Live streaming platforms like Facebook Live, YouNow, and Live.me have rolled out guest broadcasting features, where a viewer is displayed onscreen alongside the streamer and can interact with the streamer in real-time [25, 31, 53]. These prior works demonstrate that multimodal interaction in live streaming is feasible and potentially beneficial to the streaming experience. In our research, we focus on the role that multimodal communication serves in live streamed language lessons.

Multimodal Tools for Online Communication

Theories in CMC, such as the social presence theory [37] and the cues-filtered-out model [8], suggested that text-based CMC lacks nonverbal cues that are present in face-to-face communication, such as voice quality and vocal inflections, physical appearance, bodily movements, and facial expressions. This absence of nonverbal cues reduce the capacity of CMC to exchange interpersonal impressions and warmth [46]. In this work, we examine four of the most common modalities of communication apart from text: image, audio, video, and stickers. Prior research has shown that each of these modalities can enhance communication between remote users.

A number of research papers have evaluated conversation-based image retrieval and display, where a chat conversation is augmented by analyzing the keywords and automatically suggesting images related to the topic of that conversation [23, 26, 42]. Wen et al. [50] have also explored Computer-Aided Humor. Experimental results indicated that the combination of text and visual content improves emotion expression and information delivery.

The audio modality has also been investigated in various digital domains. Weisz et al. [49] found that given the options of text and audio chat, viewers preferred audio over text for talking with friends while watching online videos remotely. Geerts [11] compared audio and text chat for interactive television and showed that audio chat is considered a more natural and direct way of communication.

Prior works reveal that short videos are useful in encouraging conversation between distant users. Venolia et al. [43] found that sending reaction videos to video messages elicited authentic, engaging, and fun conversations. When sharing TV content in an online chat, Tu et al. [40] demonstrated that viewers are more responsive to lightweight content such as snapshots and video clips. Chamillionaire launched Convoz, an app where people can upload short video messages that other users can watch and respond to [36]. Similarly, the Uvii app allows users to share their opinions about a topic through videos [41]. These video-centric applications aim to elicit genuine, face-to-face conversations on social media, and reduce trolling.

In recent years, stickers have been massively integrated into online communication, particularly in mobile messaging applications [27]. Stickers are emoticons in the form of colored images. They are often animated and are different from the text-based emoticons such as :) or the in-line emojis such as 🙂 [27]. Stickers serve a variety of purposes, including emotion expression, user's self-representation, and an alternative of text [27, 54]. Animated GIFs, which are similar to animated stickers, are also shown to be the most engaging content on Tumblr [2]. In live streaming, where the conversation exchange is fast-paced, we speculate that stickers would serve as a concise and convenient way of viewer-streamer interaction.

From prior related works, we learned that richer and more expressive modalities can encourage more conversation exchange. In addition, multimedia content sharing can be useful for improving communication in social chat. In live streaming for language learning, where communication plays a significant role in successful language acquisition, the addition of multimodal cues would conceivably be valuable to the learning process.

3 METHOD

We conducted a 2-week study with one streamer and nine viewer participants to investigate the integration of multimedia tools in live streaming for language learning. We chose to study a smaller audience because we sought to understand how multimodal tools can enrich interactions, and an audience of an intimate size was more suitable for interactions on a more individual basis.

Throughout the study, the streamer taught language lessons on a live streaming application, while the viewers watched and participated through multimodal commenting. Our goal was not only to observe how *viewers* would adopt these tools for their learning but also to examine how the *streamer* would leverage multimedia interactions to support her teaching in a live stream environment. We tried as much as possible to not change the streamer's normal streaming habits. Hence, we asked the streamer to decide on the length and content of each live stream, the dates and times that these live streams happen, as well as the total number of live streams over the course of the two weeks. Our only requirement was that the

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

streamer should broadcast between two to seven live streams per week to ensure that we can gather a reasonable amount of data for analysis. In the end, the streamer broadcasted four 30-minute live streams and experimented with three different teaching strategies. During the live stream sessions, one researcher joined the stream and screen-recorded the session for later analysis. The researcher also provided technical support when needed, but otherwise did not interfere with the lessons.

Participants

We searched major live streaming and video-sharing platforms (Facebook, YouTube, Periscope, and so on) for a streamer who was familiar with language teaching using live streaming. The streamer that we found was experienced in teaching online English lessons through live streams, group video calls, and prerecorded videos. At the time of the study, she had been streaming for about a year, and was accustomed to using Facebook Live and Instagram Live for conducting live streamed lessons. In her set-up interview (see section 3.3), the streamer revealed that she used live streaming to interact with and get to know her audience, which helped her to tailor the lessons to suit their needs.

We recruited the viewers through posting on advertising websites and social networks, and reaching out to the streamer's existing students. To be eligible for the study, participants must have had experience in watching English learning live streams. Participants came from different parts of the world and had diverse backgrounds (see Table 1). The self-reported English proficiency levels ranged from intermediate to fluent. We were only able to recruit one novice speaker, who later on withdrew from the study. We speculate that novice speakers were less willing to participate because learning from live streaming required a certain level of language skills to follow along with the lesson and interact with others. Note that P8 was a fluent English speaker, but was interested in participating because she used to watch live streamed lessons to learn the British accent, and taught English online in her spare time.

System

We developed a mobile application that facilitated multimodal commenting in a live streaming system. After conducting a pilot study with the participants, we found that the system was not stable enough to ensure a smooth live streaming experience. As a result, we decided to use a combination of Facebook Live and Facebook Messenger for the formal study. Facebook Live provided reliable streaming, while Messenger included all the modalities that we would like to investigate. Although other well-known live streaming and messaging applications existed, we chose Facebook Live and Messenger because their interfaces were familiar and straightforward for general consumer use, which suited the purpose of our study.

To start a live stream, the streamer would tap on the Live button on her Facebook homepage, set the audience, enter a description for the stream, and tap the Start Live Video button to go live. The viewer participants would then receive a Facebook notification and can join the stream by tapping on the notification. The live streams were private and only visible to our participants (see Figure 1 left). While watching the stream on one device (such as a computer), each participant also engaged in a Facebook Messenger group chat on another device (such as a smartphone). Participants could send text, audio, video, image, stickers, and Facebook "like" comments (see Figure 1 right). A participant could record an audio comment by pressing and holding the microphone button while

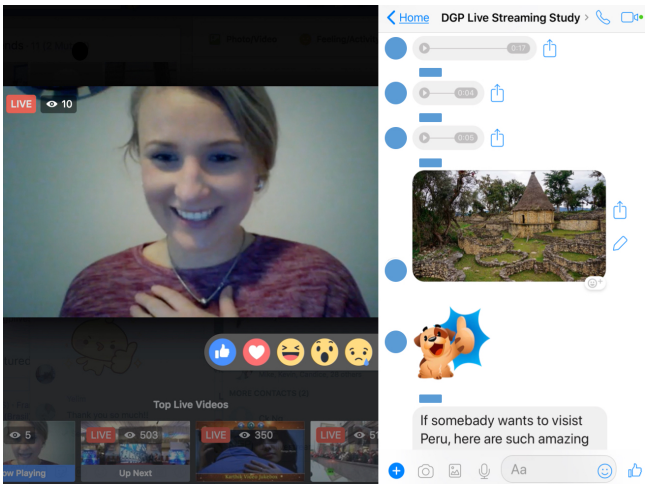

Figure 1: The language lessons were live streamed on Facebook Live (left), while the multimodal comments were sent on Facebook Messenger (right). The right figure shows three audio comments, followed by an image, a sticker, and a text comment. Participants' names and avatars are blocked for anonymity.

talking and send it by releasing the button. To send an image or video, a participant could either select existing content from her device's local storage or open the camera in-app to take a picture or video. To express herself, a participant could choose from various sets of stickers and emojis displayed in a scrollable panel. The streamer and the viewers each had access to all the comments, and could choose to view or play the comments at any time by tapping on the comment. When the streamer played an audio or video comment, all viewers could hear the audio through the streamer's live streaming device. We instructed the participants to only use Messenger (and not the Facebook Live interface) for communication.

Set-up Interviews

Before the 2-week study began, we conducted a set-up interview with each participant over video chat. The interview had three parts. First, we talked about the purpose of the study and collected informed consent. Next, we conducted a short interview and asked viewer participants about their current experiences with watching English learning live streams. For the streamer, we asked about her experience in teaching these kinds of live streamed lessons. Finally, for the pilot study, we helped each participant set up the mobile application on her device, and walked the participant through the main functionalities of the application. For the formal study using Facebook Live and Messenger, we emailed the participant instructions on how to join the stream and send multimodal comments. All participants were Facebook users before joining the study, and most of them were already

Table 2: Details of the live streamed language lessons. The lessons are denoted by L1 to L4.

	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

familiar with using Facebook Live and Messenger. We told the participant that sending and viewing comments are completely voluntary during the study.

Live Stream Sessions

Throughout the formal study, the streamer tried three different teaching strategies. For two of the four live streams, the streamer focused on *pronunciation practice*. She discussed English words or phrases that are difficult to pronounce, then asked viewers to send audio comments of themselves saying the words, and provided feedback on the viewers' pronunciations after listening to the audios. In another live stream, the streamer initiated a *conversation* between the viewers. For example, the streamer started a topic by asking a particular viewer, "what did you have for breakfast?" The viewer responded through audio or video comments and asked a question to the next viewer, and the conversation continued in this fashion. In between the conversations, the streamer corrected pronunciation, introduced new vocabulary, and talked about idioms that arose during the exercise. For the last live stream, the streamer engaged the viewers in a *picture description exercise*. She encouraged the viewers to send pictures of their country or hometown and describe the sceneries in complete sentences (see Table 2).

After each live stream ended, participants filled out a questionnaire about their experiences with the multimedia tools during the live lesson. The questionnaires served as "diaries" that allowed participants to reflect on their experiences as they occur, giving contextual insights about participants' thoughts and behaviors. We designed two questionnaires for the viewers – a short and a long questionnaire, and similarly, two for the streamer. In the short version, we asked questions related to communication, such as what the viewer mainly used each modality for and how comfortable the viewer was with sending comments of different modalities. In addition, we asked more general questions, such as examples of interesting comments and feelings about the comment modalities. This version was used for the first and third live streams. We kept this questionnaire short in order to prevent participants from losing interest and answering questions without consideration. In the long questionnaire, we preserved all

questions from the short version and incorporated additional questions about communication with the streamer and other viewers, learning experience, and engagement. We used this long version for the second and fourth live streams, which we considered as checkpoints of the study (see Table 2). We asked communication-related topics in both questionnaires because they are specific to each live stream. On the other hand, we asked about learning experience and engagement only at the checkpoints because they are higher level questions that may require observations from several live streams. The streamer's questionnaires were structured in a similar way, but the questions were tailored to the streamer's experience in the study.

Occasionally, viewers had conflicting schedules and missed one or two live streams. Whenever a viewer could not attend the live lesson, we asked the viewer to watch the archived version of the live stream to stay up-to-date on the content of the lesson. Table 2 lists the size of the audience during each of the live lessons.

Final Interviews

After the 2-week study, we performed semi-structured interviews with each participant through video chat to learn more about participants' experiences with using multimodal commenting as learning tools during live language lessons. Each interview was approximately 30 minutes long. Some questions were designed specifically for individual participants based on their questionnaire responses. For viewers, we asked about their uses for each comment modality, communication with the streamer and other viewers, how multimodal commenting affected their learning experiences, what they thought about the three different teaching strategies, their engagement, motivation, and involvement in learning, and how multimodal live streaming compared with existing live streaming. For the streamer, we asked similar questions, but from the streamer's perspective. Additionally, we inquired about modalities that were helpful for teaching English and understanding the viewers' learning progress. We probed the participants for examples of their experiences, and sought to gain deeper insight by asking how or why participants felt a certain way. At the end of the study, we compensated the streamer $96 CAD and each viewer $30 CAD for their time, and entered all participants into a draw for a $100 CAD gift card.

Data Analysis

We analyzed all interview data using an open coding approach [6]. We deliberately did not include the questionnaire responses in the open coding process. Instead, we prompted participants for clarification and more details about their questionnaire responses in the final interviews. Going through the interview transcripts, the researchers highlighted noteworthy words, sentences, or paragraphs, and created labels that summarized the data. Two researchers independently coded the first 20% of the data and met to build consensus. Then one researcher coded the remaining data, and the research team reached agreement on the codes. Finally, we used affinity diagram to group the codes and develop the themes and sub-themes.

Table 3: Frequencies of multimodal commenting during the live lessons.

	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

4 RESULTS

In total, five major themes emerged from our data analysis. We discuss the usage of the multimedia tools in the live streamed language lessons, communication and interaction between the streamer and viewers, variables that affect the use of multimodality, motivation and engagement, as well as planning and managing the lessons.

Usage of Multimedia Tools

Text. Text comments were used most frequently throughout the live stream sessions (see Table 3). Viewers used text for a variety of purposes, some of which include greetings, asking and answering questions related to the lesson, expressing opinions and thoughts, and giving feedback to the streamer. In addition, text was used to keep track of the topic at hand. In pronunciation lessons, the streamer sometimes typed out the words that she was saying, and it *"helps keep everything on track in case you miss what the teacher said"*, according to P2, *"because there are times when I'm not one hundred percent focused always on the screen"*, and text comments act *"kind of like a track that you can [use to] keep yourself going."* Viewers also used text to help others with vocabulary: *"suppose [the streamer] just says some words I don't know, especially I don't know how to spell it. Some of the students may know, so if one of them typed the word, then I can learn what the word is [and] how to spell [it]"* (P3).

Audio. Apart from text, audio was the second most used modality (see Table 3). In pronunciation lessons, viewers actively recorded audio comments and sent them to the streamer for correction. The process was iterative: viewers sent their pronunciation recordings for a word or sentence, the streamer corrected the pronunciation, then viewers attempted the pronunciation again and looked for additional

feedback. This process typically lasted for two iterations before the streamer switched to a new sentence. Due to the limited time of the live stream, not all viewers received feedback for every audio comment they sent. We discuss the limited feedback in more details in section 4.2.4. In the conversation lesson, viewers *"used audio to practice English conversation with other viewers"* (P9), or *"to practice connected speech"* (P5). Audio was also used to express feelings, answer questions, and clarify doubts. P6 mentioned, *"I sent it to express how I was feeling and to describe the pictures that other viewers and I shared in the chat group"* for the picture description practice.

Video. When practicing pronunciation and speech, some viewers sent short videos to show their mouth movements. *"It was good using the video because the teacher could see how I was pronouncing, like the movements that my mouth was doing"* (P5). *"The streamer [was] able to see how we pronounce each and every word, or you know, where we place our tongue, and how is our mouth position"* (P7). P6 sent a video to show others how the weather was in his local area at the time that the live stream was happening, *"to motivate others, to share something related to [others], [and] to feel more comfortable."* Although viewers appreciated the video modality, it was rarely used, because *"it takes more time"* (P2), and it is more prone to technical problems: *"video I think is good too, but we need to have better Internet connection to send the videos"* (P6).

Image. Image was mostly used as supporting material for refining verbal skills and starting conversations. As P5 said, *"the image, I don't think necessarily they are a direct help [with] the learning or practicing, but it helps with speaking, because we can describe the picture."* According to P7, *"[image is] useful in the way when we want to, how to say, start a topic."* Image was also a way to obtain new knowledge. As P7 pointed out, during the live stream where viewers sent pictures of their countries, *"we [were] able to get to know more about different places where everybody lives, [where] we have never been before"*, *"and I think it is good because . . . we will learn about many different cultures as well."* Although image was a valuable modality for learning, it required preparation. P2 recalled that *"[he] was scrambling to find a picture"* during the live stream. Furthermore, some images needed to be searched online. When P7 wanted to show a famous tourist attraction from her country: *"I Googled it, because I am not personally over there at the moment, so I don't have any of that picture."* The mechanism for sending online images also discouraged some viewers, as P7 continued: *"I wanted to send the link, but I find that it is not working out. So you know, I finally decided to actually download the picture and send it through my phone."*

Stickers. Stickers were used for greetings, encouragement, reacting to other participants, and expressing emotions. P2 said, *"I would use [stickers] like a thumbs up, like a happy face emoji, when somebody got something right in pronunciation of a word . . . and vice versa it started coming back too, so I felt that [it] helped a lot in the pronunciation part."* He used stickers to *"put a smile on people's face"*. P9 sent stickers to reduce awkwardness and social distance between viewers: *"in virtual class there are so many strangers and then I feel little bit awkward, I feel distant from them . . . so I think I use emoticon to praise them or encourage them, so I kind of feel closer to them."*

Multimodality as a whole. Overall, participants reported that the most useful modalities were audio and stickers. Since the streamer chose to teach materials related to speaking in all of the live stream sessions, audio was unsurprisingly the most suitable communication tool. In addition, when asked about the modalities that were most helpful for understanding the viewers' learning progress, the streamer's response was audio. Stickers were popular because they were the easiest to use, were fun and engaging, and brought viewers closer together. By contrast, image and video were used less often. There was an effort versus gain trade-off when it came to richer modalities, as the streamer remarked: *"audio and video were fantastic when it comes to feedback . . . but it was exhausting."*

Multimedia tools allowed viewers to discover areas for improvement in their language learning. For the image description lesson, P5 reflected that: *"I noticed that I wasn't so confident to talk about the pictures . . . I couldn't remember much vocabulary there, so it was great for me to see that I need to practice more on this kind of things."* As we observed, interactions through multimodal channels in live streaming may be helpful for discovering and applying new language learning techniques.

Interaction, Communication, and Connection

Interacting and connecting with others. Perhaps the most prominent perception that participants had for multimodal live streaming was that it made language learning more interactive, and *"it create[d] a lot of rooms for more communication"* (P9). Most participants felt that a sense of connection with the group was starting to blossom. For instance, P5 reported that *"when [the study] ended, I felt like we were kind of starting to feel more comfortable with each other, and starting to know more of the group, and it was kind of sad that it ended."* On the other hand, P4 commented, *"I don't think I'm connected to them, because I seldom send any comments."* While the multimedia tools served as a means of connection, it was ultimately up to the viewer whether or not to take the opportunity to interact more. Multimodality helped to bring people closer together through emotion expression, more personal interactions, and getting to know more about other

viewers' personal lives. For P9, *"I feel more closer to [other viewers] if I see their pictures or like what they are interested in."* Conversations also happened on a more personal level, and according to P3, *"this is nice because you break the ice, and just make new friends, interact with them. I had forgotten that this is live streaming."*

Desirable elements for language learning. Two leading factors that made language learning in multimodal live streaming attractive were the resemblance to real-life and personal relevance. Viewers desired to learn more practical English by mimicking real-life scenarios, and richer modalities were favorable for this purpose. P2 explained that *"as lessons progress, [viewers] are going to have full-on dialogues between each other, and this is where the video will come in handy . . . because in real-life, that's how you have to interact with people."* Similarly, P9 expressed that *"if you learn English in class, then you learn something formal, standard, . . . but if you bring up some random topic by bring[ing] up some random image, then it kind of force[s] you [to] think spontaneously in English."* Multimodal channels in live streaming also exposed viewers to different accents, which was a valuable simulation of how a language is used in real-life, especially for viewers who were accustomed to hearing the language through a standardized audio source. P5 gave an example from her personal experience: *"I remember one [job] interview was from a guy from, he was Greek, and his accent was so difficult for me [to understand], and I know that maybe if I met someone from [Greece] in these [lessons], then I would be more used to his accent because it would be more natural for me."*

Multimedia tools gave viewers more space to express themselves, which in turn made the learning more personally relevant. P5 reflected on the image description activity that: *"if [the streamer] sent us a picture, it would be a picture that a teacher sent us to describe. This was more natural, more fun because everyone could send a picture"*, and *"we could see different places with different things."* By learning from their own pictures instead of those prepared by the teacher, viewers brought snippets of their lives into the learning process. This connection to their personal lives helped viewers to contextualize what they have learned and apply the knowledge to real-life situations outside the lesson. Making the learning personally relevant is vital to remembering the material [3].

Viewer-viewer interactions. The interactions between viewers revealed some significant benefits of learning together in multimodal live streaming. In a group learning environment like this, viewers regularly compared themselves with others. Some viewers used comparison for self-evaluation. For instance, P6 shared that: *"many comments that people sent, I tried to read the comments, I tried to listen to the comments or the audios. This way I can understand better, because some of the people pronounce better than the others. This way I can realize that I'm doing well or I need to improve it."* P8 had a similar rationale: *"Listening to the audio and saying, and listening to how other people say it, of course it helps. Because you can review your own accent, you can review your own learning, basically."* Other viewers used comparison to build self-confidence. According to P4: *"I think the modality like audio, video chat, it may help to [a] certain extent . . . You may know some participant[s] who are in the same level as you. You both don't know how to speak. Suddenly, the people suddenly become confident English speaker[s], then it gives you the idea that you can actually do it like them."* When a viewer demonstrated confidence in speaking, it gave other viewers more courage to believe in themselves too.

Another benefit of having more options for communication was that viewers had more opportunities to help each other and learn together. As P7 said, *"we can learn from other people's mistakes as well as other people's strong points. So we learn together, so we grow together. It will be faster than if we grow alone."* When we asked the streamer to tell us about an experience that left a deep impression on her, she responded, *"when [the viewers] were replying to each other's audio recordings, because it shows that not only did they do [the audio recording], they took the time to listen to somebody else, and they replied."*

On the other hand, some viewers reported that their main focus was on the streamer instead of interactions with other viewers, because *"when the teacher is teaching, delivering lecture on the video, we have to be more focused on what the teacher is going to teach us . . . rather than looking at the comments sent by other participants"* (P4). *"I think direct conversation from teacher helped more, because I think other student[s] don't judge or don't evaluate my work, . . . I guess we don't really care what [other] people do, we only care about the teacher's comment"* (P9). The validity of the feedback from the streamer and the viewers were treated differently. The viewers placed more importance on the streamer's feedback because they felt that she was a more credible source to learn from. Viewers also reported listening to other viewers' responses through the streamer's playback. P7 told us that *"when the audios are [sent], usually [the streamer] will actually play it out, so usually I won't go one by one to press on the audio part to listen to it . . . because when I'm doing that, actually it obstruct[s] me from hearing what [the streamer] is going to say."* P5 also echoed, *"I wasn't playing it, because it was like double , like I would listen twice."* Participants felt that it was unnecessary to play the same audio from two different sources, and it was more important to keep up with the streamer.

Streamer-viewer interactions. Viewers' remarks about their interactions with the streamer were mostly related to feedback. They liked the instant nature of the feedback offered

by live streaming, as well as the preciseness of the feedback afforded by the multimedia tools. *"[Y]ou got a response right away from the teacher"*, said P2. *"[I]t [gives] more precise correction to our English usage"*, P9 explained. Viewers also enjoyed being able to share authentic answers. P5 reflected on her experience in a text-chat live stream prior to the study, where the streamer asked the viewers to write the answers to an exercise as text comments: *"I could just search the answer on the Internet and I could be right, but there [in multimodal live stream], [the streamer] can really see that I am the one that answered and it's a real feedback there."* The answers to pronunciation exercises could not be searched online because viewers had to send their own audio or video recordings. As a result, when the streamer gave the viewers positive feedback, it generated a more rewarding feeling.

In previous live streaming for language learning, some streamers had already incorporated asynchronous multimodal interactions into their teaching. P5 mentioned during the set-up interview that sometimes the streamer that she normally watches would leave "homework" for the viewers after the live stream. For example, the streamer would post photos of an object on her Instagram and ask her students to make a sentence using the object, or record a video of themselves talking about the object. The students would then reply to the streamer by posting the sentences and videos back on Instagram. In our study, we explored making these multimodal interactions more synchronous.

The drawback to learning language using multimedia tools in live streaming was that the number of feedback comments was finite due to the limited time that the streamer could allocate to each viewer. *"If you pronounce something wrong, the lesson kept on going, and you couldn't really fix, because by the time you decide to do audio again to fix it, they are already on something else"* (P2). For some viewers, it was difficult to keep up with the pace of the lesson. Additionally, viewers worried about taking up too much of the streamer's and other viewers' time: *"if you have the teacher one-on-one in real-life, you practice until you get it. But in live streaming, [the streamer] gave me correction and then I guess I can repeat once more and then that's probably it. There are so many other people waiting in the queue, so I cannot keep doing it"* (P9). Thus, multimodality supplied more channels for feedback, but at the same time, the amount of feedback per individual was limited.

Factors Affecting the Use of Multimodality

Concerns and constraints. Some viewers were hesitant to use a modality for a purpose that deviated from the already established purpose. For example, the audio modality was mostly used for sending pronunciations, so whenever P7 had questions or requests for the streamer, she would use text comments: *"if we have questions or need [the streamer] to repeat certain phrase or words, it's easier to reach her without confusing her with the other recordings."* In this case, the viewer felt that the purpose of the audio modality had already been defined by how it was used previously in the live stream and did not want to disrupt that "rule". The introduction of richer modalities also raised concerns about inappropriate or malicious content. This is particularly a problem for video content, and especially affects the streamer: *"if someone says something inappropriate, I don't know how to delete [the video]."*

The environment and time sometimes constrained the use of richer modalities. During the study, the live stream sessions always happened at 1 p.m. in the streamer's local time. Since participants were scattered across the globe, the stream's start time was in the afternoon for some viewers, while for others, it was in the early morning. For P3, the stream was in the afternoon: *"because the time the study was happening, I was usually at my work, so it's difficult to record."* On the other hand, P2 had just woken up for the stream, and explained that it was *"too early in the morning to see my face"*. Thus, the use of modalities was dependent on the time of the day and the viewer's surroundings.

Group size was another constraint. All participants agreed that multimodal learning in live streaming is suitable for a small number of viewers. The streamer preferred multimodal live streaming *"in a group, maximum 10 attendees, give or take that one or two don't attend."* From the viewer's point of view: *"it works well with small groups, not with big groups, because in small groups, the streamer can listen or read all the comments and give a comment too."* When the group size becomes large, viewers may lose interest in the lesson because they cannot get enough attention and feedback. Furthermore, it would be more difficult for the streamer to decide on which comments to view or listen to. Group size and multimodality also affect the organization of comments. In existing text-chat live streams, comments become unmanageable for a larger audience [33]. Having multiple modalities for communication undoubtedly creates additional challenges for managing comments.

Identity and self-image. Viewers displayed varying levels of willingness to send comments of richer modalities. One determining factor pertained to feeling uncomfortable with revealing personal identity online. For P9, *"I don't think it is unuseful, but for video I feel very awkward to put my face and then share my video to people."* Interestingly, identity exchange was not expected to be reciprocal, as P9 continued, *"for me I don't like to show my identity, but if someone [is] willing to show their identities and then opens up, and I kind of can see where they are living, how they look like, [then] of course it feels more familiar than just talking through the text . . . I feel more close if I get to know their face and voice and*

how they talk." Revealing identities brought people closer together because it showed the realness of the viewers who were behind screens, but it was not required to be an equal relationship. Majority influence was another prominent factor in deciding the use of richer modalities. P2 recounted, *"I have pictures of me outside, but nobody was sending pictures of themselves outside, they were just sending the outdoors. So I felt kind of awkward to send a picture of myself outdoors . . . ."* Participants' self-consciousness and the desire to fit in made them hesitant to act differently from others.

Motivation, Engagement, and Involvement

Interaction with others was the most apparent motivation for learning language using multimedia tools in live streaming. In particular, viewers' enhanced communication with the streamer pushed them forward, as P7 shared: *" if the streamer responds to the comments or audio or images, actually that is a great encouragement because it helps us to want to do more."* Interacting and getting to know other students also heightened the sense of community: *"I felt listened, I felt being part of a group"* (P8). Another motivation for viewers was getting acknowledged for the progress that they were making: *"sometimes I felt excited to continue participating because it's good when somebody makes some comments about what I am doing"* (P6). Multimodality is helpful for this kind of supportive learning because more channels are available to showcase the viewer's skills and track the viewer's progress.

Unsurprisingly, most participants stressed that having fun while learning kept them motivated and engaged. P5 said, *"it wasn't like a normal or formal class, we could like have fun while we were studying."* P7 neatly summarized the reason why the "fun" factor was crucial to learning: *"when it is fun, you tend to do more."* P6 told us that *"[i]t's more interesting with multimodalities than just text, because just text is kind of boring"*. By making dull lessons more fun, multimodality effectively created a more captivating experience that differentiated it from existing text-chat live streaming lessons. However, it is important to note that due to the short duration of the 2-week study, viewers' elevated interest in the lesson could be caused by the novelty effect.

Throughout the study, viewers played different roles in the live streams depending on their level of involvement. Most participants were active learners who eagerly participated in the English practices. In contrast, the role of lurker also emerged, but the rationale behind lurking was unanticipated: *"the reason for not sending any comments is that I give priority to other participants. I think they need more help than me . . . At least I don't mind just listening to what the teacher is going to say in the video"* (P4). The participant yielded the right of talking and getting corrections to other viewers. Several participants filled the role of encouraging or motivating others, either because they felt empathetic: *"I know how it feels like to try to learn another language and have to struggle"* (P2), or they tried to lighten the mood which, according to P9: *"that relaxes me and then makes me enjoy the class more."*

Lesson Planning and Management

During the final interview, the streamer emphasized reducing the "teacher talking time", which is a teaching style that involved less talking by the teacher and encouraged more communication between the students [1]. *"When I'm [teaching], I like to reduce teacher talking time, and this [multimodality] really enabled me to reduce teacher talking time."* The viewers were sharing the streamer's burden of maintaining the conversation in the live stream and generating new conversation, by sending different modalities of comments. The streamer pointed out that out of the three teaching strategies, the pronunciation lessons required the most effort from her because the *"teacher talking time is high, and also correcting and explanation [needs effort]."* However, for lessons that required less teacher talking time, such as the conversation lesson, the streamer had to prepare an alternative plan: *"I had some talking points . . . so if it [conversation] doesn't work, then I'll teach them connected speech."* With a small audience, it was hard to predict how participative the viewers will be and when the conversation will end.

Some viewers speculated that a live streamed lesson is more manageable for the streamer than other online learning methods such as group video calls. Although the streamer had no control over the comments that viewers sent, she had the authority to decide which comment to look at or listen to. P5 recalled, *"I think she listened to all of them, but she had control of it."* The streamer, on the other hand, had psychological pressure to listen to or view all comments: *"I just feel like, there's six of them [in the live stream], I can be a decent person and listen to all of their comments."* Because the multimedia tools allowed for more personal interactions, the streamer felt a greater sense of responsibility towards viewers who put effort into sending the comments.

5 DISCUSSION

Supporting Large Scale Audiences

We learned through the study that the most important contributions of multimodality to the learning process were the instant feedback and the increased interactions between the participants involved in the live stream. However, participants mentioned that multimodal learning in live streaming is suitable for a small audience. With a larger audience, the streamer would not have time to interact individually with each viewer. Giving students individual attention is not only a challenge in live streaming, but more broadly, it is a problem associated with teaching in general. As the number of students increases, the attention and time that the teacher can distribute to each student will naturally decrease.

In order to support larger audiences for multimodal interactions in live streaming, one possible approach is to assign moderators to live streamed lessons. Moderators are viewers who are given privileges to perform administrative tasks for the streamer during a stream. They are common in large scale live streams and play an important role in the stream. According to research on Twitch streaming communities, moderators engage viewers and promote participation; they often greet viewers, answer questions, and connect personally with the viewers [15]. In a live stream for learning, moderators can serve a role similar to classroom TAs. When a large audience makes it difficult for the streamer to split her focus and engage with viewers, moderators can fill in the gap, helping the streamer answer questions and promote active learning.

Another potential approach to support multimodal learning in live streaming at scale is to split the audience into smaller interaction groups. Prior research has shown that small-group video discussions are a promising way to promote interactions and learning for large scale global classes [4]. Likewise in a large scale live stream, the audience can be split into smaller comment groups managed by a moderator to participate in conversations and ask questions. The streamer can flexibly divide the audience to start small-group interactions, or merge the audience to refocus on the streamer's teaching materials. Naturally, this approach would require an effective way to manage and group comments, which we discuss in more details in section 6.1.

Comparing Live Streaming to Other Synchronous CMC

Live streaming, when used as a language learning platform, presented some fundamental differences compared to other synchronous CMC platforms like video conferencing and network broadcasted talks. In video conferencing, the teacher has the same access as the students to all audio and visual modalities. The communication is synchronous between the teacher and students, allowing multiple participating parties to talk at the same time. Because the teacher has little control over who is talking, asking a question requires turn-taking, and audience size must be limited. On the other hand, the comment system in multimodal live streaming is more asynchronous since the streamer does not have to listen to an audio or watch a video as soon as a viewer sends it. Viewers do not need to take turns to submit comments, and anyone can view or respond to the comments. The viewer's expectation for a response from the streamer is also lower in a large scale live stream because of the huge volume of comments.

In network broadcasted talks, the learning that takes place is generally more formal and standard than in live streams. The speaker is usually prepared with slides to teach, and the content of the talk is mostly fixed [21]. In a live stream, the audience is more empowered to influence the content of the stream. Whereas the speaker for a network broadcasted talk is typically stationed at a fixed location such as a desktop workstation or a lecture room, live streaming is more flexible; streamers can start a live stream almost anywhere: at home, at work, on the bus, and so on. The informal environment in which the live stream is broadcasted creates more opportunities for casual conversations, which is vital for building language skills. Similar to video conferencing, interactions in network broadcasted talks require turn-taking. Students must wait in a queue to ask questions, and questions are addressed mostly by the teacher. In live streaming, peer interactions are more viable because all comments are simultaneously available to all viewers, but at the same time, interactions can quickly become messy.

Through exploring multimodal live streaming for language learning, we discovered that different learning platforms are suitable for different groups sizes and formats of learning. Multimedia tools in live streaming allow for more participation from the viewers, while also giving the streamer some control over the lesson. Although the modalities are not equally important or relevant to the learning, each of them contributes in their own ways to the communicative and interactive learning process.

Positioning the Studied Context

Our study investigated language learning, which is a specific type of knowledge sharing content seen in live streams. In fact, a diverse range of knowledge sharing content exists in live streaming; examples include academic learning such as mathematics and psychology [32], live coding (e.g. [19]), physical activities like yoga [34], and life skills such as cooking [32, 39]. We recognize that our findings may not necessarily generalize to these other types of knowledge sharing content. Furthermore, we studied an audience of an intimate size, and our results may not apply to other scales of interactions, such as one-on-one online tutoring or massive online courses. Finally, our participants had diverse cultural, technological, and demographic backgrounds. These factors may have had an impact on participants' use of multimodal tools during the study.

6 DESIGN IMPLICATIONS

Based on our analysis of the study, we discuss some implications for the design of multimodal live streaming systems to support language learning, which can be applied more broadly to learning in general.

Supporting Effective Comment Display

Multimodality brought about increased participation and engagement in live streams for language learning. However, it also introduced serious challenges, one of which is the

cluttering of comments. Live streaming systems enhanced with multimedia tools should effectively manage the display of multimodal comments, especially with a larger number of viewers. A possible solution may be providing the option to filter the comments, for example, by modality, topic, or viewer name. Another approach is employing message organization techniques such as conversation threading, which are widely used by messaging applications and collaboration tools like Slack [48]. The system could also extend existing solutions for text-chat based live streaming systems to multimodal live streaming, such as strategically limiting the comments that a viewer sees [33].

Distributing Responsibility

In the study, we saw that the viewers mainly focused on interactions with the streamer. With a small audience, the streamer felt more pressure to address the viewers' comments. Viewing comments of richer modalities also required more mental effort. Multimodal live streaming services should incorporate tools to ease the streamer's burden. For example, as the study indicated, viewers regarded the streamer's feedback as more credible compared to feedback from other viewers, which resulted in comments been primarily directed toward the streamer. The system could encourage viewers to answer each other's questions and employ a mechanism to increase the credibility of a viewer's response, similar to reputation points on Q&A sites such as Stack Overflow [20]. Distributing the responsibility of answering questions among the viewers reduces the streamer's cognitive load while promoting a more supportive and participatory learning environment.

Sharing the Streamer's View

Viewers revealed that they listened to their peers' audio comments mostly when the streamer played the recordings as opposed to by themselves since they were mainly fixated on the streamer. Other viewers' comments were helpful mainly because a viewer can learn from their mistakes when the streamer gives feedback on the comments. Platforms that support multimodal live streaming should consider displaying the multimodal comment that the streamer is currently viewing to all viewers, especially for the richer modalities. We argue that doing so would provide viewers with more context for what the streamer is saying, and improve the shared learning experience.

7 LIMITATIONS

We recognize that the study had limitations. First, we had a small sample size and a relatively short study period. However, we believe that the 2-week duration still allowed us to sufficiently reveal significant use cases for multimodal tools while maintaining the participants' interests in the study. Additionally, the live streaming service and the multimedia tools were on two separate applications, and could not coordinate with each other. When a viewer was recording an audio comment on Facebook Messenger, the application could not automatically mute the streamer's audio from Facebook Live. As a result, some audio comments were less audible because of the background noise from the live stream. Fortunately, viewers quickly overcame this problem and shared their solutions with others. They either manually muted the live stream's audio or put on earphones to isolate the sounds. Despite the inconvenience, the main intention of the study, which was to explore the effect of various multimodal tools in live streamed language lessons, was achieved.

8 CONCLUSION

Using an empirical approach, we investigated how incorporating multimedia tools in live streaming for language learning affected communication, engagement, and interactive experiences between the teacher and students. Although the use of multimedia has long been studied in language learning and Computer-Mediated Communication, our work is the first longitudinal in-the-wild study to explore multimodality usage for language learning in the live streaming medium. Multimodality brought about more natural and personal interactions, generated a higher level of engagement, and further promoted active learning; but at the same time, challenges such as supporting larger scale audiences remain to be solved.

ACKNOWLEDGMENTS

We would like to thank all participants of our study, the reviewers for their valuable feedback, and members of the DGP lab for their support.

REFERENCES

[1] Richard L Allwright. 1984. The importance of interaction in classroom language learning. *Applied linguistics* 5, 2 (1984), 156–171.

[2] Saeideh Bakhshi, David A. Shamma, Lyndon Kennedy, Yale Song, Paloma de Juan, and Joseph 'Jofish' Kaye. 2016. Fast, Cheap, and Good: Why Animated GIFs Engage Us. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 575–586. https://doi.org/10.1145/2858036.2858532

[3] Sara Bernard. 2010. Science Shows Making Lessons Relevant Really Matters. Retrieved April 19, 2018 from https://www.edutopia.org/neuroscience-brain-based-learning-relevance-improves-engagement

[4] Julia Cambre, Chinmay Kulkarni, Michael S. Bernstein, and Scott R. Klemmer. 2014. Talkabout: Small-group Discussions in Massive Global Classes. In *Proceedings of the First ACM Conference on Learning @ Scale Conference (L@S '14)*. ACM, New York, NY, USA, 161–162. https://doi.org/10.1145/2556325.2567859

[5] Karen Church and Rodrigo de Oliveira. 2013. What's Up with Whatsapp?: Comparing Mobile Instant Messaging Behaviors with Traditional SMS. In *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI '13)*. ACM, New York, NY, USA, 352–361. https://doi.org/10.1145/2493190.2493225

[6] Juliet Corbin, Anselm Strauss, et al. 2008. Basics of qualitative research: Techniques and procedures for developing grounded theory. *Thousand Oaks* (2008).

[7] Gabriel Culbertson, Solace Shen, Malte Jung, and Erik Andersen. 2017. Facilitating Development of Pragmatic Competence Through a Voice-driven Video Learning Interface. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1431–1440. https://doi.org/10.1145/3025453.3025805

[8] Mary J Culnan and M Lynne Markus. 1987. Information technologies. (1987).

[9] Miguel Angel Farias, Katica Obilinovic, and Roxana Orrego. 2011. Engaging multimodal learning and second/foreign language education in dialogue. *Trabalhos em linguística aplicada* 50, 1 (2011), 133–151.

[10] Kim Flaherty. 2016. Diary Studies: Understanding Long-Term User Behavior and Experiences. Retrieved January 7, 2018 from https://www.nngroup.com/articles/diary-studies/

[11] David Geerts. 2006. Comparing Voice Chat and Text Chat in a Communication Tool for Interactive Television. In *Proceedings of the 4th Nordic Conference on Human-computer Interaction: Changing Roles (NordiCHI '06)*. ACM, New York, NY, USA, 461–464. https://doi.org/10.1145/1182475.1182537

[12] Abbas Pourhossein Gilakjani, Hairul Nizam Ismail, and Seyedeh Masoumeh Ahmadi. 2011. The Effect of Multimodal Learning Models on Language Teaching and Learning. *Theory & Practice in Language Studies* 1, 10 (2011). https://doi.org/10.4304/tpls.1.10.1321-1327

[13] Susan Goldin-Meadow. 2014. Widening the lens: what the manual modality reveals about language, learning and cognition. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 369, 1651 (2014). https://doi.org/10.1098/rstb.2013.0295

[14] Oliver L. Haimson and John C. Tang. 2017. What Makes Live Events Engaging on Facebook Live, Periscope, and Snapchat. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 48–60. https://doi.org/10.1145/3025453.3025642

[15] William A. Hamilton, Oliver Garretson, and Andruid Kerne. 2014. Streaming on Twitch: Fostering Participatory Communities of Play Within Live Mixed Media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 1315–1324. https://doi.org/10.1145/2556288.2557048

[16] William A. Hamilton, John Tang, Gina Venolia, Kori Inkpen, Jakob Zillner, and Derek Huang. 2016. Rivulet: Exploring Participation in Live Events Through Multi-Stream Experiences. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video (TVX '16)*. ACM, New York, NY, USA, 31–42. https://doi.org/10.1145/2932206.2932211

[17] Regine Hampel and Eric Baber. 2003. Using internet-based audiographic and video conferencing for language teaching and learning. *Language learning online: Towards best practice* (2003), 171–191.

[18] Yuki Hayashi, Aoi Sugimoto, and Kazuhisa Seta. 2017. Accessible Multimodal-interaction Platform for Computer-supported Collaborative Learning System. In *Proceedings of the 11th International Conference on Ubiquitous Information Management and Communication (IMCOM '17)*. ACM, New York, NY, USA, Article 82, 4 pages. https://doi.org/10.1145/3022227.3022308

[19] Suz Hinton. 2017. Lessons from my first year of live coding on Twitch. Retrieved December 30, 2018 from https://medium.freecodecamp.org/lessons-from-my-first-year-of-live-coding-on-twitch-41a32e2f41c1

[20] Stack Exchange Inc. 2018. Stack Overflow. Retrieved February 11, 2018 from https://stackoverflow.com/

[21] Ellen A. Isaacs, Trevor Morris, and Thomas K. Rodriguez. 1994. A Forum for Supporting Interactive Presentations to Distributed Audiences. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work (CSCW '94)*. ACM, New York, NY, USA, 405–416. https://doi.org/10.1145/192844.193060

[22] Gavin Jancke, Jonathan Grudin, and Anoop Gupta. 2000. Presenting to Local and Remote Audiences: Design and Use of the TELEP System. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '00)*. ACM, New York, NY, USA, 384–391. https://doi.org/10.1145/332040.332461

[23] Yu Jiang, Jing Liu, and Hanqing Lu. 2016. Chat with illustration. *Multimedia Systems* 22, 1 (01 Feb 2016), 5–16. https://doi.org/10.1007/s00530-014-0371-3

[24] Kyung Je Jo, John Joon Young Chung, and Chung Juho Kim. 2017. Exprgram: A Video-based Language Learning Interface Powered by Learnersourced Video Annotations. (2017).

[25] Liz Jostes. 2017. How to Invite a Guest to Facebook Live from your iPhone. Retrieved January 6, 2018 from https://www.elirose.com/2017/04/invite-guest-facebook-live/

[26] Joon-Gyum Kim, Chia-Wei Wu, Alvin Chiang, JeongGil Ko, and Sung-Ju Lee. 2016. A Picture is Worth a Thousand Words: Improving Mobile Messaging with Real-time Autonomous Image Suggestion. In *Proceedings of the 17th International Workshop on Mobile Computing Systems and Applications (HotMobile '16)*. ACM, New York, NY, USA, 51–56. https://doi.org/10.1145/2873587.2873602

[27] Joon Young Lee, Nahi Hong, Soomin Kim, Jonghwan Oh, and Joonhwan Lee. 2016. Smiley Face: Why We Use Emoticon Stickers in Mobile Messaging. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '16)*. ACM, New York, NY, USA, 760–766. https://doi.org/10.1145/2957265.2961858

[28] Pascal Lessel, Alexander Vielhauer, and Antonio Krüger. 2017. Expanding Video Game Live-Streams with Enhanced Communication Channels: A Case Study. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1571–1576. https://doi.org/10.1145/3025453.3025708

[29] Mike Levy. 2009. Technologies in Use for Second Language Learning. *The Modern Language Journal* 93, s1 (2009), 769–782. https://doi.org/10.1111/j.1540-4781.2009.00972.x

[30] Ulf Liszkowski. 2014. Two sources of meaning in infant communication: preceding action contexts and act-accompanying characteristics. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 369, 1651 (2014). https://doi.org/10.1098/rstb.2013.0294

[31] Live.me. 2016. 'Beam' Lets You Add Viewers Into Your Broadcasts. Retrieved January 6, 2018 from https://medium.com/live-me/beam-lets-you-add-viewers-into-your-broadcasts-4331b0c6701b

[32] Zhicong Lu, Haijun Xia, Seongkook Heo, and Daniel Wigdor. 2018. You Watch, You Give, and You Engage: A Study of Live Streaming Practices in China. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 466, 13 pages. https://doi.org/10.1145/3173574.3174040

[33] Matthew K. Miller, John C. Tang, Gina Venolia, Gerard Wilkinson, and Kori Inkpen. 2017. Conversational Chat Circles: Being All Here Without Having to Hear It All. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 2394–2404. https://doi.org/10.1145/3025453.3025621

[34] Reese Muntean, Carman Neustaedter, and Kate Hennessy. 2015. Synchronous Yoga and Meditation over Distance Using Video Chat. In *Proceedings of the 41st Graphics Interface Conference (GI '15)*. Canadian Information Processing Society, Toronto, Ont., Canada, Canada, 187–194. http://dl.acm.org/citation.cfm?id=2788890.2788923

[35] English Like A Native. 2017. LEARN Phrasal Verbs: The Complete List - #1 | Live English Lesson. Video. Retrieved January 7, 2018 from https://www.youtube.com/watch?v=VVJa-seiQek

[36] Katie Roof. 2018. Chamillionaire is a presentation genius, has a new app. Retrieved March 22, 2018 from https://techcrunch.com/2018/02/11/chamillionaire-is-a-presentation-genius-has-a-new-app/

[37] John Short, Ederyn Williams, and Bruce Christie. 1976. The social psychology of telecommunications. (1976).

[38] Brendan Sinclair. 2018. Outpost wants to mix games with reality TV. Retrieved March 22, 2018 from https://www.gamesindustry.biz/articles/2018-03-09-outpost-wants-to-mix-games-with-reality-tv

[39] John C. Tang, Gina Venolia, and Kori M. Inkpen. 2016. Meerkat and Periscope: I Stream, You Stream, Apps Stream for Live Streams. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 4770–4780. https://doi.org/10.1145/2858036.2858374

[40] Pei-Yun Tu, Mei-Ling Chen, Chi-Lan Yang, and Hao-Chuan Wang. 2016. Co-Viewing Room: Mobile TV Content Sharing in Social Chat. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 1615–1621. https://doi.org/10.1145/2851581.2892476

[41] Uvii. 2018. Uvii. Retrieved March 22, 2018 from https://www.uviiapp.com

[42] Jeroen Vanattenhoven, Christof van Nimwegen, Matthias Strobbe, Olivier Van Laere, and Bart Dhoedt. 2010. Enriching Audio-visual Chat with Conversation-based Image Retrieval and Display. In *Proceedings of the 18th ACM International Conference on Multimedia (MM '10)*. ACM, New York, NY, USA, 1051–1054. https://doi.org/10.1145/1873951.1874147

[43] Gina Venolia, John C. Tang, and Kori Inkpen. 2015. SeeSaw: I See You Saw My Video Message. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '15)*. ACM, New York, NY, USA, 244–253. https://doi.org/10.1145/2785830.2785847

[44] Gabriella Vigliocco, Pamela Perniss, and David Vinson. 2014. Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 369, 1651 (2014). https://doi.org/10.1098/rstb.2013.0292

[45] VIPKID. 2018. VIPKID. Retrieved September 16, 2018 from https://t.vipkid.com.cn/

[46] Joseph B Walther, Tracy Loh, and Laura Granka. 2005. Let me count the ways: The interchange of verbal and nonverbal cues in computer-mediated and face-to-face affinity. *Journal of language and social psychology* 24, 1 (2005), 36–65.

[47] Mark Warschauer and Deborah Healey. 1998. Computers and language learning: An overview. *Language teaching* 31, 2 (1998), 57–71. https://doi.org/10.1017/S0261444800012970

[48] Social Media Week. 2017. Slack Announces 'Threads' for More Organized Conversations. Retrieved March 1, 2018 from https://socialmediaweek.org/blog/2017/01/slack-threads-organized-conversations/

[49] Justin D. Weisz and Sara Kiesler. 2008. How Text and Audio Chat Change the Online Video Experience. In *Proceedings of the 1st International Conference on Designing Interactive User Experiences for TV and Video (UXTV '08)*. ACM, New York, NY, USA, 9–18. https://doi.org/10.1145/1453805.1453809

[50] Miaomiao Wen, Nancy Baym, Omer Tamuz, Jaime Teevan, Susan T Dumais, and Adam Kalai. 2015. OMG UR Funny! Computer-Aided Humor with an Application to Chat.. In *ICCC*. 86–93.

[51] Yuli Yeh and Chai wei Wang. 2003. Effects of Multimedia Vocabulary Annotations and Learning Styles on Vocabulary Learning. *CALICO Journal* 21, 1 (2003), 131–144. http://www.jstor.org/stable/24149484

[52] Dongwook Yoon. 2015. Enriching Online Classroom Communication with Collaborative Multi-Modal Annotations. In *Adjunct Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15 Adjunct)*. ACM, New York, NY, USA, 21–24. https://doi.org/10.1145/2815585.2815591

[53] YouNow. 2015. Introducing Guest Broadcasting. Retrieved January 6, 2018 from http://blog.younow.com/post/128644352834/introducing-guest-broadcasting

[54] Rui Zhou, Jasmine Hentschel, and Neha Kumar. 2017. Goodbye Text, Hello Emoji: Mobile Communication on WeChat in China. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 748–759. https://doi.org/10.1145/3025453.3025800

[55] Yeshuang Zhu, Yuntao Wang, Chun Yu, Shaoyun Shi, Yankai Zhang, Shuang He, Peijun Zhao, Xiaojuan Ma, and Yuanchun Shi. 2017. ViVo: Video-Augmented Dictionary for Vocabulary Learning. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 5568–5579. https://doi.org/10.1145/3025453.3025779