Relevance and turn-taking in cross-modal computer-mediated communication: a case study of online live streaming platforms

Qingyun Catherine Wang

A thesis submitted in partial fulfillment of the requirements
for the degree of Bachelor of Arts in Linguistics

Swarthmore College
Spring 2023

I. INTRODUCTION

With the prolific rise of instant messaging and video conferencing in the past decade, it is undeniable that a large portion of modern linguistic discourse occurs not through face-to-face interaction, but through digital platforms, such as via phone, text messaging, or perhaps most commonly, the Internet. As a result, the study of computer-mediated communication (CMC), though a recent and emerging subfiled in linguistics, is gaining interest and popularity both within and outside of academia, as both synchronous and asynchronous computer-mediated communication become an increasingly necessary and integral part of digital interaction and thus a central aspect of modern life. In the age of user-generated content and social media platforms, the internet is a ripe source of natural language and conversation, making it an incredible resource of accessible data for linguists, especially discourse analysts interested in interactive behavior and its change over time and across mediums. However, interactive behavior is also subject to incredible variability even within small groups, as well as rapid change that makes these phenomena difficult to examine and quantify.

Early forms of computer-mediated communication (CMC) primarily occurred on asynchronous, text-based platforms, such as emails and web forums. However, modern online discourse is increasingly multi-user, multi-channel, and multimodal, making CMC environments far more rich and complex than they were a mere few years ago. For example, video and text chat are increasingly used in tandem in online social, business, and educational settings, with web conferencing and live streaming platforms also gaining massive traction worldwide. Such settings involve a diverse set of communication applications and interfaces whose features and affordances allow for and shape a diverse variety of interactional behaviors. Nonetheless, native users of digital and Internet communication platforms will recognize common behaviors and interactive phenomena that remain constant across different applications and settings. Such constants underlie the core phenomena that we may consider to be computer-mediated communication.

Despite the potential for insight into the nature of communication presented by digital multimodal discourse mediums, the rapid pace of technological changes often outpace the opportunity for academic investigation. As a result, modern changes and innovations in CMC, its userbase, and its pragmatic patterns still hold great potential for linguistic insight. The relative newness of online multimodal environments means that how they fit into traditional frameworks of conversational analysis remains an open question. Past forays into CMC research have observed that technological features of the medium (e.g. instant messaging or voice chat) allow for robust, multi-user, and fast-paced interaction, and often times result in disruptions or complete decays of conversational aspects such as sequential relevancy and turn management, especially in multi-user CMC. Such evaluations lead to the argument that CMC lacks interactional coherence entirely (Herring 2006); nonetheless, complex CMC environments continue to grow in user base and popularity, suggesting that, at least to their native users, some manner of conversational coherence must be upheld in order for any fruitful interaction to occur, though they may perhaps hold loosened—or entirely different—expectations for relevance and turn-taking management. As such, it remains a question of interest how and for what goals users navigate interaction in digital environments.

This thesis will focus in particular on a digital environment that unites the 'incoherent' nature of text chat with cross-modal interactions with voice and video conversation, while simultaneously influenced by a unique set of social structures and rules that are seemingly unique to the online environment. Specifically, I will look at live streams, a form of communication in which a broadcast host communicates to an audience through a real-time video broadcast on the Internet, as a multimodal discourse environment. The streamer and audience operate in two distinct channels, one using voice and video broadcasting, and one using text chat. Both independent and cross-modal interactions exhibited between the two will be shown to exhibit behavior different than that observed in single-channel, single-medium CMC. Through case studies of live streaming environments online, I will aim to create a broad framework of the discourse environment of live streams as well as make an effort to more

accurately describe the influence of technological and social features on participants' expectations when it comes to interaction, relevance, and turn-taking.

Ultimately, the goal of this study is to provide a descriptive account of interaction in live streams and propose how the cross-modal nature of the stream influences discourse behavior with regard to the aforementioned discourse features. Through a review of the literature, I will establish the capability of live streams to behave generally like conversation, in which multiple technological mediums facilitate dialogic communication between the streamer and the audience as well as among the audience. I will then through my case studies observe how such communication can occur across the very different mediums of text and speech and what strategies participants use to facilitate it, and what rules and assumptions govern the structure of this conversation that, by its nature, is multifaceted and often disjointed. By performing observational case studies of multiple live streams and comparing cross-modal interaction to existing literature on text-only and video-mediated CMC, this thesis will seek to explore the way that the discourse environment of live streams is shaped by technological, cognitive, and social factors, and broadly argue for the relationship between these factors when it comes to linguistic analysis of online communication.

II. BACKGROUND

I will begin by providing a review of the existing literature on computer-mediated communication, specifically of interactional behavior on text-only and voice and video platforms. The goal of this section is to not only provide relevant context for the objectives of this study, but also to provide a baseline for comparison against which I will evaluate the influence of multi-modality on discourse behavior. I will first provide an overview of existing research that takes discourse approaches to CMC, followed by necessary background on the technological and social environment of live streaming.

2.1 The discourse & pragmatics of computer-mediated communication

Given the unique form of communication presented by both text- and video-based CMC, it is unsurprising that scholarly attention has turned to describing this communication through the lens of discourse and pragmatics. In the following subsections, I provide an overview of established behavioral norms first in text-only then in video-mediated CMC, and of existing literature that not only describes such behavioral norms, but also posit potential explanations that account for the differences between them and the practices traditionally observed in face-to-face communication.

2.1.1 A brief history of the field

The study of computer-mediated communication as a linguistic field has expanded alongside the rise of user-generated content on the Internet. Though the term encompasses a wide range of natural language interaction that can occur through digital devices, modern study of computer-mediated communication heavily focuses on language use on the Internet, perhaps in large part due to the abundance of easily accessible linguistic data available there. As interactive and social website programming proliferated in the late 1990s and early 2000s, the participative or social web, also known as

Web 2.0, brought user-generated content such as blog posts and social media to the forefront. This type of content began to rapidly outpace static content, such as articles and corporate webpages (Herring 2013; Murugesan 2007). In such an age of user-generated content, and buffeted by the rising influence of social media, the Internet is now a ripe source of naturally occurring language and conversation, and thus a source of interest for linguists seeking organic instances of user interaction and behavior. At the same time, however, such behavior is subject to extreme variability even within small groups, as well as rapid change that makes its phenomena often difficult to empirically define. As a result, empirical linguistic analysis of the pragmatics and discourse of CMC remains difficult.

Early forays into linguistic analysis of CMC were largely concerned with classifying CMC as a whole in comparison to face-to-face communication and written communication, with a focus on superficial structural features, which to many seemed to characterize the mode as a whole. Early studies of online forums and text chats highlighted such distinctive properties, such as unique lexical items, abbreviations and acronyms (e.g. "LOL"), emoticons (typed features meant to indicate emotion, such as "**:**)"), and non-standard punctuation (Werry 1996). However, as the rapid nature of language change online has caused many of these structural features to fall out of use, attention has instead turned away from the lexicon and surface features of CMC, and rather to the underlying discourse structure it produces.

Ferrara, Brunner, & Whittemore (1991) suggest that interactive text-only CMC might constitute an emergent register, which they call Interactive Written Discourse, a hybrid register showing features both of written and spoken language. While some recognized the nature of CMC to behave like spoken language, many considered it to be a 'lean' form of communication, in some ways suffering from the lack of the multifacetedness of face-to-face interaction, in which language use is frequently accompanied by intonation, prosody, and physical gestures that orient participants in a shared space, as well as giving rise to more ambiguity (Crystal 2001). However, as digital communication has grown to become an increasingly dominant form of communication for many, such negative and simplistic evaluations of CMC have given way to further characterizations of CMC that suggest that it behaves similarly to spoken

language in ways it previously did not, or was generally assumed not to. Baron (1984) predicted that the "loss of communicative nuances that comes with decreased feedback" puts increasing pressure on users to make use of any means possible to ensure they are being understood, making CMC a potential site of rapid language change. Discourse studies of online language use in the 1990s and early 2000s, focusing on forums, email chains, and instant messaging chats populated by American English speakers[1], found that pragmatic phenomena generally associated with conversation and not with writing could be observed in these environments, such as etiquette and interaction management (e.g. Rintel & Pittam 1997; Markman 2005). The growth of the Internet in other countries in the mid-1990s further inspired cross-linguistic research into language use online (Herring, Stein, & Virtanen 2013).

With the development of instant messaging and social media platforms, text-based CMC has not only begun to more and more closely approximate synchronous conversation, but also began to exert an influence on communication and social interaction. With the proliferation of such platforms, much language use on the Internet has been dubbed 'quasi-synchronous', as posted messages are available synchronously, but the message production process is available only to the participant producing the message, and thus a delay is formed between the production and the transmission of the message (Garcia & Jacobs 1999). With this developed an interest in the pragmatics of CMC, with topics such as relevance and speech acts in different modes and genres of CMC. Many studies have observed that the written record left by text-based communication coupled with the rapid, quasi-synchronous nature of instant messaging allow for many communicators to engage in multiple threads of conversation at once (Simpson 2013). Due to the ability to read and respond to past messages, multiple ongoing conversations may be sustained at once, and the maxims that are generally considered to universally regulate conversation may be relaxed or not even present. Moreover, the 'lean' medium of written text encourages participants to utilize creative textual means to articulate tone or gestures, further fostering language play and the

[1] As social media becomes more widespread, with commercial platforms seeing increasing global success, it becomes difficult to claim that users of a particular platform are members of a particular cultural, national, or linguistic demographic without potentially invasive investigation. Though study of specific dialects and cultures as they interact through digital and social media is entirely possible, I do not claim that the data investigated in this study is local to a specific region.

development of a dynamic and complex discourse environment in many online platforms (Herring, Stein, & Virtanen 2013).

In the past decade, and especially after the widespread impact of the COVID-19 pandemic, voice and video communication has become increasingly popular and necessary, underscoring the question of how video-mediated communication interacts alongside the 'quasi-synchronous' nature of text-based CMC. Though many studies of CMC have historically observed its usage in creative and social interactions, the dominance of digital communication in formal settings, especially education, post-2019 has inspired further curiosity into characterizing interaction and discourse in CMC. With much CMC research, the goal has been to examine the extent to which human interaction and the linguistic maxims by which it operates are influenced by mediation via technology, and in what way these influences surface in participant choices and behavior. As such, there remains a need for descriptive studies of CMC environments that feature not only text-based interaction, but audio and visual mediums as well. In the following sections, I will first address the established literature on the pragmatics of text-based synchronous CMC before continuing on to the ways in which voice and video CMC environments are structurally similar to and different from text-based environments.

2.1.2 Text-based CMC

The earliest forays into the pragmatics of CMC investigated asynchronous, text-based conversation that was characteristic of early digital communication. For instance, email was one of the first forms of CMC that received scholarly attention, especially in the field of text linguistics. The creative use of capitalization, spelling, and punctuation to add emphasis or emulate prosody has been observed since the early 2000s (Durscheid & Frehner 2013). Moreover, as the rate of response becomes faster, emails have been observed to behave less like letters, with users forgoing typical greetings and farewells, and increasingly using anaphoric expressions to refer to previously presented information, especially when participants are viewing the interaction as conversational (Severinson Eklundh 2010). This led the

way for pragmatic analyses of text-based CMC, especially as more rapid and conversational forms of digital communication, such as instant messaging, grew in popularity.

As text-based digital communication methods expanded, so too did academic interest in the ways that it was similar to and different from spoken communication. Many studies of CMC pragmatics have focused on topics like relevance, turn-taking, and topic flow in multi-participant internet relay chats (IRC), an early text-based chat system for instant messaging, in which group communication occurs in channels (designated forums for discussion), while one-on-one conversation can also occur in private messages. Though IRC usage has declined since the early 2000s, many modern instant messaging platforms have a similar user interface, facilitating similar conversational behaviors. Paolillo (2002) found in a study of multi-participant IRC that overall turn lengths tended to be extremely short, between three and six words per turn, with turn adjacency frequently interrupted by new topics, as well as new members joining the channel. Participation in IRC is fluid, with users logging on and off as they wish, without the need for conversational orientation speech acts commonly observed in face to face interactions (e.g. greetings or goodbyes). This form of text chat structure remains common across the internet, including in the text chat of live streams, which I will examine in this paper. Thus, it is important to establish the existing characterization of the pragmatics of multi-user, quasi-synchronous text chat before turning to the interaction this text chat may have with a voice and video broadcast.

In conversation analysis, conversation is often characterized as a set of interactional exchanges, where utterances are expected to be complete and adjacent to one another (Sacks & Schegloff 1973). However, highly active IRC rarely display such behavior. Even in one-on-one conversations, the quasi-synchronous nature of IRC means that common adjacency pairs are frequently incomplete or appear out of order. For instance, the following example from Simpson (2013) shows a case where a response can appear before a question, due to the delay needed to type out the question utterance and the second participant providing the information before the first participant has finished asking.

```
<MichaelC> Good evening Ying. How are things?
```

```
<Ying-Lan> Not so good.
<Ying-Lan> I took a test this morning.
<MichaelC> What's wrong?
```

In multi-participant conversations, the rate of such disruptions are magnified, giving rise to the notion of the perceived lack of coherence in CMC. Herring (2006) observed a high degree of overlapping exchanges and topic decay in multi-participant IRC, and described highly active IRC as 'interactionally incoherent'. Generally, coherence is defined as the logic and semantic relation between units of text, such as sentences or propositions; in multi-participant text-based CMC, immediately adjacent units of text rarely hold this property. Multiple threads of conversation frequently simultaneously spawn within a single channel (Danet 2001), and participants make contributions to past conversational topics or respond to messages that are not immediately adjacent to their contribution without the need to explicitly signal that they are doing so, often in the middle of other ongoing conversation. However, such studies rejected the characterization of text-based CMC as fully incoherent, arguing that "if CMC were seriously incoherent, users would not flock to the Internet so enthusiastically" (Herring 2006). Clearly, users are capable of managing at least some aspect of the disrupted adjacency and disrupted interaction that occurs in multi-user text chat; however, many conversational maxims previously considered universal are predictably and systematically violated. The rest of this section will cover some of the existing literature on quasi-synchronous text-based CMC and discuss the ways in which it has been observed to behave differently than face-to-face conversation, as well as remaining open questions about the explanation of this behavior and if it is inherent to CMC.

Relevance in Interaction

When it comes to traditional pragmatic phenomena, relevance is central to the carrying out of conversation and discourse. Grice (1975) posits relevance as one of the four maxims of cooperation: one of many necessary expectations that are required for conversation to be productive and coherent. Sperber & Wilson (1986), in their foundational work in Relevance Theory, suggest that conversations tend

towards optimal relevance because cognition seeks maximal relevance, not necessarily because individuals are consciously adhering to an expectation such as Grice's Cooperative Principle. Relevance Theory also argues that relevance is a precursor to coherence in a cognitive way; even an uncooperative, completely self-interested speaker would still need to maintain relevance in a conversation, or their audience will have no reason to pay attention at all. As a result, conversational participants tend to hold a presumption of optimal relevance, assuming that any given statement is not only relevant given the current context, but is also the most relevant contribution to be made at that time. In such a way, we may consider the question of 'relevance' in discourse to not be a binary one (i.e. 'is this utterance relevant or not?') but rather a scale of prioritizing multiple relevant inputs.

In these traditional definitions of relevance in conversation, logically-related utterances are expected to occur sequentially in time. Thus, conversation exhibits *sequential relevance*, where conversation is locally structured as a series of adjacency pairs, where each utterance is not only locally relevant to the previous turn, but globally relevant to the current conversational topic, except when intentionally bringing up (or attempting to bring up) a new topic of conversation. An utterance that is perceived as unrelated to the current sequence of conversation is likely to be perceived as violating or intentionally flouting the maxim of relevance.

However, text-based CMC has been observed to not uphold this expectation of sequential relevance. First, quasi-synchronous text chat frequently involves disrupted adjacency of utterances that would otherwise be logically consistent, which is common when multiple users are transmitting messages at the same time, given the delay caused by the need to type out textual messages. The following example from Paolillo (2011) is a commonly-cited example of the typical lack of interactional coherence in IRC:

```
<ashna> hi jatt
*** Signoff: puja (EOF From client)
<Dave-G> kally i was only joking around
<Jatt> ashna: hello?
<kally> dave-g it was funny
<ashna> how are u jatt
<LUCKMAN> ssa all
```

```
<Dave-G> kally you da woman!
<Jatt> ashna: do we know eachother?. I'm ok how are you
*** LUCKMAN has left channel #PUNJAB
*** LUCKMAN has joined channel #punjab
<kally> dave-g good stuff:)
<Jatt> kally: so hows school life, life in geneal, love
life, family life?
<ashna> jatt no we don't know each other, i fine
<Jatt> ashna: where r ya from?
```

The above sequence of messages may appear incoherent when read strictly sequentially, but most users nonetheless would not find the exchange difficult to comprehend, and can parse the topics with relatively little ambiguity. This suggests a relaxed expectation of sequential relevance in synchronous CMC (Herring 2013), possibly challenging the universality of optimal relevance.

Secondly, the nature of text-based conversation to leave a written record allows users to view and respond to messages that are not sequentially immediate, without the cognitive burden of mentally keeping track of several instances of past utterances. In fact, modern instant messaging platforms allow users to receive messages even when not online, such that new users joining a text chat in which a conversation is ongoing can 'backread'[2] a past or ongoing conversation and make a contribution relevant to a past statement that is not salient in the current state of the conversation. This differs significantly from face-to-face conversation, in which new participants have no such means to catch up on missed conversation without interrupting the current conversation by asking another participant to relay past information. Thus, it follows that expectations for new participants to only make immediately relevant contributions is significantly relaxed in text chat, leading to less strict expectations for sequential relevance in general.

Third, even in the case of turns intended to be related, expectations for local relevance remain weak. It is typical of utterances to only be 'weakly' related to the utterance to which they are ostensibly responding to, in the sense that even only marginally relevant responses are frequently accepted as normal

[2] Wikitionary users define this Internet slang term as "to catch up on an ongoing conversation, by reading previous portions one was not present for". The general existence of meta-commentary regarding backreading in online chats hints at the ubiquitousness of such behavior for native users.

aspects of discourse (Herring 1997). As such, online text-based conversation in forums and group chats is often characterized by rapid topic development and death, with tangential observations and weakly related contributions being far more frequent (Lambiase 2010); Herring (1999) proposes that the structural and technological nature of multi-user text chat accounts for this by entangling multiple threads of conversation, making it more cognitively difficult to keep track of long sequences of exchanges for many topics at once. However, recent research has found that the technological medium itself may not be the only factor, as social setting (e.g. formal vs. informal) is likely to affect the amount to which off-topic contributions are accepted (see Stormer-Galley & Martinson 2009).

Nonetheless, Relevance Theory remains important in the analysis of synchronous CMC, as users clearly must cognitively recognize contributions that are and are not relevant, and reject irrelevant contributions where they are made. Moreover, in an environment where multiple contributions can be generated and transmitted simultaneously, users must make relevant contributions in order to receive attention and response from other users. As a result, participants must in some sense compete for the control of the conversational floor (or floors); how participants systematically gain (or fail to gain) floor control remains a question.

Floor control & topic flow

The conversational floor is constructed very differently in CMC than in face-to-face conversation. In face-to-face conversation, only one user is generally expected to speak at one time, although interruptions and overlaps may occur. Therefore, there is frequently no distinction between the floor and the turn, even in multi-party conversations; however, Edelsky (1981), in a study of turn development and participation in face-to-face group discussions, argues that simply speaking is not enough to gain control of the conversational floor. The floor, as Edelsky defines it, is dependent on how participants evaluate the topic, function, or interaction going on in a given conversation. Speakers and listeners work together in maintaining it, in that floor-holding turns must be ratified by other participants through acknowledgment (e.g. verbal or non-verbal backchannel) or response. Edelsky identified two ‘types’ of conversational

floor: a ‘linear’ type characterized by a singular thematic focus and ‘orderly’ one-at-a-time turn-taking, and a ‘collaborative’ type characterized by overlapping speech, shorter turns, and broader turn distribution. Herring (2010) conducts a parallel study of Edelsky’s study focusing on group discussion on the Internet rather than face-to-face, concluding that social factors such as gender and power impact the development of ‘collaborative’ vs. ‘linear’ floors in CMC, calling for an explanation that rests on social, rather than technological, elements.

Though asynchronous text-based CMC (such as mailing lists and forums) may allow for the development of conversations that contain a singular, ‘linear’ conversational floor, in the case of quasi-synchronous CMC, especially in multi-participant scenarios, the aforementioned nature of conversational adjacency pairs to be interleaved and interrupt sequential adjacency results in the development of multiple conversational threads and contributes to a fast rate of topic development and decay. It is difficult to empirically quantify the rate of topic development, even in a text-based medium, as it involves evaluating semantic shifts and their local and global relevance. Nonetheless, CMC is consistently characterized both in literature and in popular perception as ‘off-topic’ and subject to large and fast-paced topic shifts. Herring (1999) suggests that since quasi-synchronous CMC involves a lack of feedback, participants cannot know what other participants are writing when producing their own utterances, thus creating branching responses to any given utterance. As a result, each possible statement produces multiple competing directions for the conversation to take; participants are therefore, in some sense, competing for their contributions to gain control of the conversational floor(s).

In many online environments, such as forums and chat room channels dedicated to a specific topic, implicit (and sometimes explicit) rules enforce on-topicness, preventing completely unrelated and ‘random’ messages from being sent, but such rules are frequently difficult to enforce. This is particularly interesting in the case of live streams; though live streams are publicly broadcast, unlike open chats and forums, interaction in live streams still occurs on a certain broadcaster’s channel, where the broadcaster

and moderators[3] may enforce channel rules and remove participants who violate them. Often times, these rules, in addition to social pressure, prevent participants from engaging in discussion irrelevant to the live stream, such as engaging in inappropriate or harmful conversations, spamming irrelevant messages, or promoting one's own content or goods on another user's broadcast. However, within the acceptable realm of topics, there is often no set topic that the stream will necessarily be restricted to. As a result, it is still reasonable to hypothesize observing a high rate of topic decay in live streams.

Timing and turn-taking

As previously discussed, synchronous text-based CMC frequently exhibits disrupted adjacency and overlapping exchanges, making timing and turn-taking highly different than face-to-face conversation. In face-to-face conversation, participants must actively yield the channel itself to one another, as multiple people speaking over one another does not make for productive conversation for any participant. However, in text-based CMC, due to its quasi-synchronous nature, multiple participants can construct and send their messages at one time, with no need to yield time or space within the channel to one another. Cherny (1999) states about synchronous CMC that "given that there is no competition for the channel per se, but rather competition for attention or control of the discourse, notions of shared or collaborative floor seem to be more helpful than the standard turn-taking literature." Garcia & Jacobs (1999) point out that since users can neither interrupt and prevent another participant from speaking, nor control the exact placement of their message relative to what other participants will post, the use of traditional face-to-face turn-taking techniques, as identified in Sacks, Schegloff, & Jefferson (1974), are not productive in exchanging talk turns in quasi-synchronous CMC systems. The lack of simultaneous feedback is problematic for many face-to-face turn-taking techniques; unlike in face-to-face

[3] Like many forums and text-based chat spaces allow for community moderators who may delete messages and ban users, many live streaming platforms allow broadcast hosts to appointment chat moderators who may similarly manage the text chat. Note the distinction between chat moderator, an appointed role on the platform, and conversational/discourse moderator, a role that a conversational participant assumes when they are in a position to manage interaction in a given discourse setting. For instance, a teacher in an online classroom would act as a discourse moderator (and potentially a chat moderator as well, depending on their duties).

communication, in quasi-synchronous CMC, a given participant cannot see what another participant is typing until the other participant hits 'send'. For example, in oral conversation, a speaker might select a next speaker with the first part of an adjacency pair, using an address term or physical gesture to indicate a specific individual who should respond; however, this caused issues as the selected next speaker may be typing a different message, reading past messages, or simply not actively engaged in the conversation, leaving room for other participants to jump in.

Moreover, face-to-face conversations typically involve a stream of ongoing backchannel that involves both physical gestures and phatic verbal expressions, allowing a participant to provide responses during another participant's turn without taking the turn (Yngve 1970). While there is evidence to suggest that conversational backchannel does exist in CMC, especially facilitated by technological features such as the ability to 'react' to another poster's message with a response or emoticon without interrupting the current conversation, the lack of conversational backchannel makes interruptions and thus interrupted adjacency far more common in CMC. At the same time, users can easily view past messages despite interruptions, meaning that disrupted sequential adjacency poses a significantly lower risk to conversational coherence in text-based CMC than it does when speaking face-to-face.

However, this is not to say that users do not engage in active turn-taking strategies. Explicit conversational hand-offs and self-selection are frequent especially in multi-user conversations, especially when a participant declares the need to make a contribution in advance or explicitly holds their turn knowing that it will take them time to produce another utterance due to the delayed transmission of typed text. Participants might explicitly ask others to wait for them to finish before moving on to another topic, despite the fact that the written nature of quasi-synchronous CMC means that no 'interruption' is truly possible[4]. This suggests that to some extent, turn-based conversation is a valid framework through which

[4] In my experience in multi-participant instant messaging on the Internet, highly active multi-participant conversations are likely to induce such common turn-taking practices. For instance, a user might make a contribution to a certain topic before indicating that they are about to make a contribution to another topic, or simply make an utterance expressing that they have not finished making their contribution. Though this remains highly anecdotal, further study may be warranted to examine how turn-taking practices in text-only CMC mirror face-to-face conversation when in highly active multi-participant interactions.

to consider quasi-synchronous CMC, even though it may appear at first glance that face-to-face turn-taking strategies are not in employment.

2.1.3 Voice & video

Though video-mediated communication is a subset of CMC, it appears intuitive that, due to the spoken nature of interaction, such communication will more greatly resemble face-to-face interaction than text-based CMC does. Unlike text-based CMC, which leaves a written record, video-based communication usually does not, meaning that the relaxed expectation of relevance and ability for multiple topics of conversation to co-occur tends not to be present. Video-mediated communication in fact requires even more cases of interactional management, such as greeting and response adjacency pairs, to establish mutual orientation when speakers are not physically present in the same space (Brandt & Jenks 2013). Similar to voice-only digital communication, such as telephone calls, participants anticipate identification and greeting sequences, which may be as simple as a ‘hello’ or ‘are you there?’.

In multi-participant voice or video based communication, turn-taking and timing becomes even more important, as opposed to the previously discussed text-based CMC. Unlike in text-based CMC, overlapping talk can occur in video-based communication, and is counterproductive to productive communication. However, unlike in face-to-face communication, technological factors such as video delay or simply the lack of physical gestures oriented to a shared space prevent traditional turn-taking practices from taking place. Instead, the strategic use of pauses and repair and correction strategies are necessary to manage inter-turn coherence in voice based CMC (Jenks 2009). In formal settings, such as online classrooms, an authority such as an instructor may be expected to moderate turn-taking, calling upon participants (such as students) to make contributions at given intervals (Earnshaw 2017). Participants may also utilize features of the technological medium, such as turning off their microphone to relinquish a turn, and turn on their microphone to signify an intent to make an oral contribution, thus self-selecting as the next speaker.

Given the newness of the medium, studies on video-mediated communication remain fairly sparse, and frequently focus on how the video communication often overlaps with simultaneous written text, such as in the case of conversational video-mediated communication over platforms such as Skype, Google Hangouts, or Zoom. For example, Earnshaw (2017) studied virtual classroom environments and found that both students and educators, who act as moderators of classroom discussion, often make use of dual-channel methods (specifically, turning to the use of chat in addition to video) for the purpose of conversational repair, such as when a statement is misunderstood or there is a technological issue. Users frequently modulate between both channels as needed for such purposes.

2.1.4 Interaction across mediums

As mentioned previously, many conversational forms of video-mediated communication, such as a Zoom conference or Skype call, allow users to simultaneously communicate through multiple channels, often a speaking channel and a written one. Thus, a system of simultaneous verbal exchanges co-exists with a quasi-synchronous one, often involving 'mode-switching', in which users frequently switch back and forth between speaking and writing within the same conversation (Sindoni 2013). Such digital conferences create a virtual space, often referred to as a *room*, in which each user is allotted space in the speaking and video channel as well as the ability to participate via the chat channel.

Such discourse environments, which are both multi-channel (maintaining a separate spoken and written channel) as well as multi-modal (utilizing both video and text mediums), often exhibit specific discourse maxims surrounding the different channels and modes. Not only do participants use the text channel when necessary for clarification or repair when the spoken channel poses a communicative issue, participants also choose to use the text channel to intentionally make certain contributions to the conversation that may be regarded as less salient or less involved (Rosenbaun et al. 2016).

Research also shows that the written chat is frequently used to hold parallel floors in multi-user conversations (Sindoni 2014). Unlike text chat, spoken conversation is generally not productive when

there are many interruptions and overlaps; thus, the speaking floor can usually only be held by one participant. Nonetheless, the existence of text chat gives users the opportunity to provide responses to spoken statements without interrupting a spoken response, thus allowing an alternative conversational topic to be raised on a parallel floor.

2.2 The technological environment: multi-modal & multi-channel

In this section, I will begin to characterize the exact environment that I aim to describe through this study: online live streams. In this study, I consider the discourse environment of online live streams to be a factor of both technological and social features. Thus, I will first characterize what those technological features are, as well as provide background to the actual technological structure, layout, and logistics of the live streaming environment, before characterizing broader social features in the next section.

2.2.1 Defining multi-modality and cross-modality

In discourse and conversation analysis, the term ‘multimodal’ generally refers to a framework of conversation analysis that takes into account the multiple forms of communication that occur in a single interaction. In addition to the words spoken, interlocutors make use of and take into account gestures, eye contact, and movement in physical space. The multimodal framework of discourse considers all of these as discourse acts that may influence the ongoing conversation. In this paper, I refer to multi-modality in a much stricter sense: the literal use of multiple mediums in one interactional setting. Thus, I refer to the live streaming environment, as described above, as a multimodal discourse environment, as the streamer and audience operate in two distinct channels, one using voice and video broadcasting, and one using text chat. This differs from a discourse environment such as a forum or chat room, in which all participants are using the same mode of communication.

However, unlike many commonly-studied video-mediated communication forms, such as video conferencing rooms, live streaming does not allow for 'mode-switching' by the majority of participants. Audience members are limited to using chat only, and thus must compete for acknowledgment and recognition of their contributions in the speaking channel. On the other hand, the streamer has the ability not only to speak and write, but also to present or display whatever they wish in their stream, be it a game screen, videos, text, or anything else that can be projected on a computer screen, providing a further medium to communicate (though this will not be a focus in this study).

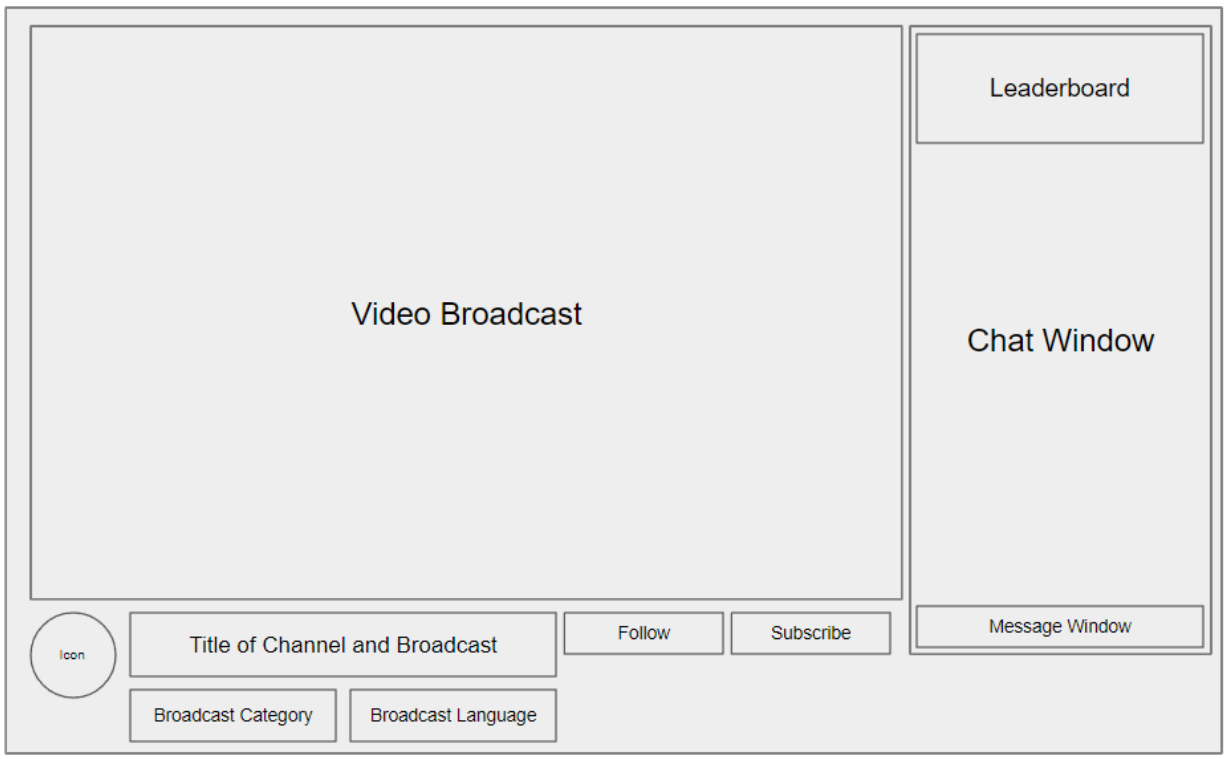
Nonetheless, communication undeniably occurs across the spoken and written mediums in a live stream, and this communication is necessarily impacted both by the technological factors of each medium as well as the nature of the discourse environment. In this paper, I will refer to such interactions as 'cross-modal', when dialogic utterances occur across the spoken and written mediums. For example, in a virtual classroom, this may be when an instructor asks a question over video chat and students respond in the text channel. In a live stream, I will define this as interaction between the streamer, who is using the spoken channel, and the audience, who is using the text channel. Such interactions are likely to be influenced by social factors, especially when considering the streamer's role in the stream environment. However, we must also examine how the nature of each medium impacts the conversational timing, expectations of relevancy, and other pragmatic and discourse phenomena observed in these interactions, as well as how these may differ from the norms established in research of 'mode-switching' forms.

2.2.2 The technological environment of live streams

A live stream is a form of media that is simultaneously recorded and broadcast over the Internet. Specifically, pre-recorded broadcasts, such as vlogs or video-on-demand, may be streamed online, but these are not live streams, as the recording does not occur in real time with the broadcast. If a viewer is to make a comment on such a video, they will not be able to do so, nor will the broadcaster ever see it, until after the video has been completely finished and uploaded. In live streams, since recording and

broadcasting occurs simultaneously, audience members often expect (and are expected to) interact directly with the live streamers. All popular commercial live streaming platforms, such as YouTube, Twitch, Facebook Live, Instagram Live, bilibili, and Kuaishou, allow audience members to directly participate with live streamers in some way by commenting using text chat. The hosts of the live stream are able to read these messages in real time and respond in the broadcast.

Fig. 1: Schematic representation of Twitch's user interface



Most commercial live streaming platforms consist of a video broadcast accompanied by a simultaneous chat that may appear next to, below, or overlaid on top of the video stream. Audience members interact within the chat, which may boast features common to Internet instant messaging systems, such as customized usernames, standard and customized emojis (images embedded in text chat upon command), and reply functionality. Some live stream platforms, such as Twitch, involve specific features for highlighting text, such as Twitch Superchat, which allows channel subscribers to pay to have

their message highlighted. Other interactive events, such as polls or predictions, may occur and appear overlaid above the chat window.

A specific live streamer's broadcasts are usually hosted on that broadcaster's 'channel'. A 'leaderboard' showing the users who have donated or gifted the most to the channel recently may sometimes appear above the text chat window. Options to follow the channel (i.e. receive notifications for the broadcast) or subscribe to the channel (i.e. pay a monthly subscription to Twitch to support the channel or receive certain benefits) are also highlighted below the broadcast window.

While the broadcasters typically occupy the spoken channel, they may also choose to type in the chat for a variety of purposes, and may also appoint moderators to help manage the stream and community. Bots and AI, which may provide information on command such as rules, relevant links, or stream information, are also present in the chat and can interact with users. A user may interact with a bot by sending a specific command, and the bot can respond by displaying information in the chat that is visible to all participants.

2.3 The social environment: the live stream

The last decade has seen a boom in the market and community for live streaming, with an increasing diversity of activities such as video gaming, drawing, music, cooking, eating, etc. being broadcast. Commercial platforms, most notably Twitch, YouTube, bilibili, and Facebook Live, have seen their user bases expand massively during this time. While some platforms, such as Twitch, characterize themselves as explicitly catering towards video gamers and entertainers[5], other platforms cater to a broader demographic, hosting lifestyle, commercial, and even educational live streams. The expectations for formality and etiquette are thus likely to highly differ between platforms and channels; regardless, even these informal environments enforce a set of strict social rules, either implicitly or explicitly declared on the live streamer's home page. Moreover, the relationship between the live stream host and

[5] Twitch's company website, as of 2023, states that "this is the home for creators streaming video games, music, sports, and everything else they love with magnetic authenticity", and their partnerships and branding historically cater primarily towards video games and e-sports.

their audience further underscores the host's role as both a performer and moderator of discourse, commanding the audience and establishing the common ground upon which the social interaction rests.

2.3.1 Micro-celebrity and the streamer—audience relationship

The phenomenon of micro-celebrity (Senft 2008) is undoubtedly tied to the rise of social media and live streaming. The participatory culture of social media encourages the majority of online content to be collaborative and/or derivational, from the popularization of viral memes to the creation of fanworks to the activity of watching live streams itself. In such a connected era, the behavior of existing online is inherently tied to maintaining and marketing one's online identity, often as though it were a branded good, with the expectation that others are doing the same (Senft 2012). Unlike 'mainstream' celebrities, the premise of micro-celebrity rests upon the perceived closeness with their fans, or in other words, the performance of intimacy. The participatory nature of social media and live streaming encourages performers to engage directly with audience members and capitalize upon fan engagement to achieve social and economic success. Thus, speech acts must be carefully curated to allow for maximum engagement. Traditionally, such performances include seemingly mundane discussions of everyday life, giving fans the illusion of "backstage" access and close intimacy to their idols (Marwick & boyd 2011). The frequency with which fans may interact directly with celebrities through social media and live streams further encourages live streamers to offer fans an illusion of exclusive closeness, thus incentivizing financial and social support of the live streamer.

Moreover, with large, public live streams, especially by established streamers who market themselves through a connection with their followers and community, comes an inherent parasocial relationship between presenter and audience. In media and communication studies, parasocial relationships (Horton & Wohl 1956) refer to a one-sided relationship that fans form with celebrities or idols resulting from one-way intimate interactions; while fans perceive themselves as knowing the celebrity closely due to being aware of the activities and even lifestyle of the celebrity, there is no actual

close relationship as the celebrity does not know the fan personally. Xu et al. (2021) found in a study of Chinese college students that participants perceived their relationships with micro-celebrities as more reciprocal than that with 'mainstream' celebrities, due to the ability to communicate frequently with micro-celebrities on the Internet. This suggests that the Internet and its affordances such as live streaming make the branding of the self and the performance of authenticity with fans much more accessible, such that almost anyone could do it.

This relationship forms the basis of social interaction in the live stream setting. Although in smaller, more 'intimate' streams, frequent audience members may be familiar with one another, in larger streams, it is highly unlikely that most audience members will know one another, making live streams a very different environment than a video conference or virtual classroom, in which audience members are likely to have a coworker or classmate relationship with one another. Thus, the common ground is established through each audience member's relationship with the live streamer and the aforementioned perception of intimacy and community with them. Audience members are motivated to become central participants in the conversation and thus the community, but this participation is directly controlled by the broadcast host, whose choices to respond to chat statements or ignore them (or impose stricter regulations such as banning a user or deleting their messages) affect audience members' abilities to participate in the discourse environment.

2.4 The discourse environment of live streams

Though existing literature on the discourse of live streams is sparse compared to other forms of CMC, especially text-only synchronous CMC, some studies have investigated live streams as a discourse environment and attempted to characterize the behavioral norms native to live streams relative to other forms of CMC as well as to spoken conversation. Hamilton et al. (2014) broadly categorized small to medium sized live streams as conversation-like, while large streams behaved like a presentation, with the many members in the audience causing the chat to appear like a waterfall of text with no meaningful

dialogic conversation occurring. Recktenwald (2018) characterizes live streams, especially video gaming streams, of being primarily monological, with streamers and viewers both producing monological moves that discuss the game and goings-on of the stream without having a direct recipient. However, as Recktenwald points out, dialogical communication between streamer and audience, within the audience, and between the streamer and others in the video game also frequently occur.

It is difficult to empirically chart discourse acts on a quantitative scale such as audience size. Nonetheless, recognizing that there may exist a certain number of active participants above which meaningful discourse is difficult or even impossible is beneficial when it comes to characterizing discourse and interaction in live stream settings, as we want to avoid broadly generalizing live streams as largely non-interactive when that is only the case when the number of active participants is high to a certain degree. In fact, even from a largely simplistic and anecdotal viewing of many "medium-sized" live streams, it is quite evident that in many live streams, streamers and users expect to interact with one another through dialogic speech acts.

Recktenwald (2018) describes live streaming, like quasi-synchronous text-based CMC, as producing an environment with multiple conversational floors, in which the spoken floor has a prioritized status over the textual floor. The streamer, being the only participant allowed to speak due to the technological structure of the live stream, has a monopoly over the spoken floor, and as follows the ability to moderate and control the flow of discourse.

Unlike the speaking channel, the text channel can sustain multiple conversational floors, due to the aforementioned nature of text chat to leave a record and allow for multiple threads of conversation to co-occur. As a result, audience members not only participate in conversation in the textual floor(s), but also compete for floor control both in the chat and, in many cases, in the speaking channel by attempting to gain notice by the streamer. As such, authority is highly relevant in the discourse structure of live streams, with even the most active participants in the audience prevented from ever fully gaining floor control without the recognition and permission of the broadcaster. This fixed configuration of modes (who

is allowed to speak vs. who is allowed to write), as well as the fixed configuration of roles (e.g. live streamer, moderator, chatter) directly structures the participation framework of a live stream.

2.5 The motivations for this study

The relative newness of video-mediated communication and video broadcasting over the Internet leaves many unanswered questions about the ways in which communication and conversation is affected by technological medium. At the same time, it is highly important to consider the social impact of the Internet and the communities it forms. Thus, this study will aim to consider both these factors in a quantitative analysis of interaction in the live stream setting. It seeks to examine claims about conversational practices and sequential organization in CMC in multi-channel settings, to extend existing frameworks about CMC discourse, examine the ways in which conversational expectations change to accommodate multiple mediums and channels of discourse, and ultimately evaluate such frameworks and their productiveness in the study of CMC and discourse as a whole.

Given that the average Internet user now engages in more frequent and more complex instances of computer-mediated communication than even a mere few years ago, it is more important than ever to move towards a comprehensive analysis of CMC that takes into consideration the increasing ways that CMC platforms and behaviors become more robust and perhaps, in some ways, more similar to face-to-face communication. In particular, in light of the COVID-19 pandemic, CMC is increasingly integrated into daily life, with virtual settings becoming a highly accessible and common option for education, business, and entertainment. It therefore is especially prudent for linguists, especially conversation and discourse analysts, to focus on emerging trends in computer-mediated discourse and the ways in which they will influence interaction in social, educational, and business settings. Thus, descriptively characterizing the nature of interaction in multimodal settings becomes increasingly necessary for the goal of designing better technological tools or educational methods that facilitate meaningful and productive conversation through digital mediums.

III. METHODOLOGY

This section will provide an overview of the discourse analysis strategies employed in this study, as well as the details of how case studies were selected and data retrieved and processed.

3.1 Methods for research on CMC

Herring (2004) notes that while the Internet and the rising popularity of user-generated content has caused a boom in research on human behavior, creating a massive, permanent, and searchable database of language and interaction, it remains difficult for researchers, and linguists in particular, to make generalizations about trends of online behavior that remain empirically grounded. The newness of Internet research tends towards a tendency to broadly generalize labels for online phenomena, as well as focusing on specific, often superficial, features that may appear to separate online discourse from in-person talk.

However, it is undeniable that much online interaction occurs through discourse and conversation. Even more static examples such as blog posts are inherently recipient-designed, argue Meredith & Potter (2013), who propose the usefulness of conversation analysis (CA) methods in designing electronic interactions. They point out that research that focuses on specific practices or forms, such as abbreviations, emojis, and non-standard spellings, will take such practices out of context if they are not oriented in the conversational and interactional context in which they are used. Given that conversation analysis aims to capture how social interactions are managed in everyday practice, it is important that data collection ensures that participants are behaving the way they normally would; at the same time, it is preferable to use screen recording softwares and capture what is live for a given participant at a given time, including partial messages as the participant types them into their chatbox. This is necessary to gain a complete sense of timing, turn-taking, and orientation within a quasi-synchronous environment. Of course, gathering such data comes with a host of practical and ethical concerns regarding participants' privacy and consent.

Though it is true that the timing and production of partial messages is of importance to the conversation analysis of CMC, it is still possible to conduct in-depth analysis of the ways in which interactions are managed and occur without recording each individual's screen, and instead looking at transcriptions of the conversation as it is posted publicly and counting the presence of discourse phenomena of interest. Here we will turn back to Herring (2004) and its characterization of a discourse analysis approach to CMC, which Herring calls computer-mediated discourse analysis (CMDA), focusing on empirical analysis of verbal interaction through the sampling of text. CMDA makes three primary assumptions:

1. Discourse exhibits recurrent patterns (Goffman 1959)
2. Discourse involves speaker choices
3. CMC discourse is partially (though not necessarily completely) influenced by the technological features of the medium

In keeping with such assumptions, Herring proposes using language-focused content analysis as a basis for discourse analysis of CMC. Specifically, the researcher should define a digital environment and perform a sampling of the verbal behaviors that occur in that environment, whether through random sampling, chronologically, or by phenomenon.

This makes CMDA a useful lens by which to approach the type of online environment-based study that I aim to accomplish. By understanding utterances and conversational turns as the core linguistic unit of interaction and sampling such utterances to examine for a strictly defined set of discourse phenomena, it is possible to build up a framework of interaction and discursive behavior in a specific environment. Therefore, I will perform a CMDA analysis on case studies of live streaming environments, specifically performing a phenomenon-based sampling of text chat messages that implicitly or explicitly modulate relevancy, as well as the relevant (both conversationally and temporally) interactions going on in the live stream.

3.2 Analyzing live streams online

Now that I have established the value of a discourse-centric approach to CMC research, I will provide my methodology for carrying out computer-mediated discourse analysis on live streams and justify the case studies and sampling methods I chose.

3.2.1 Selecting appropriate case studies

To select an adequate set of broadcasts for the purposes of this study, many factors have to be taken into account, both for the purposes of feasibility of research as well as appropriateness for the research topic. Firstly, the platform of the broadcast is important. Although computer-mediated discourse analysis does not operate under the assumption that the features of the technological platform necessarily define the discourse features observed on it, there are still many reasons why the platform matters. First of all, many platforms may be much more popular in one part of the world than another, so we should consider the language and geographic distribution of users. While Chinese live streaming app Douyu is one of the largest live streaming platforms in the world, boasting more monthly active users than Twitch, the decision to perform this study on English speakers is incompatible with the primarily Chinese-speaking userbase. Second, different online platforms often have differing expectations of formality and politeness due to the userbase and marketing; for instance, although Twitch is a massive platform that features an extensive range of broadcast topics, many of which may be educational or business oriented, it is undeniable that the platform's branding as one that caters towards video game, e-sports, and entertainment live streamers impacts the expectations for formality on the platform.

Another important factor is the size of the live stream. Existing analysis on the discourse of live streams argues that the size of the audience strongly influences the nature of interaction among users. Hamilton (2014) characterizes 'small to medium' sized streams as acting more like conversations (in the discourse sense), while 'large' streams act more like presentations. Of course, such analysis is necessarily

incomplete, as it is difficult to create a quantitative scale of exactly how discourse changes based on the number of people present in the audience, or provide any definitive cutoff that labels a stream as 'small', 'medium', or 'large'. Nonetheless, we ought to consider that it may be unproductive to analyze fairly large live streams for conversational or dialogical speech acts. As audience sizes grow, it becomes increasingly unfeasible for any given participant to expect to carry on a conversational exchange with another, and much more likely to contain monological contributions, such as cheering for the streamer, spamming emojis, or otherwise one-way reactions. At the same time, we want to balance this both with the amount of text chat content and interaction that exists in the chat. We do not want a live stream in which very little engagement is happening, as our goal is to track the modulation of relevancy as multiple topics intersect with one another across the different mediums. Thus, we need to look for broadcasts whose size and 'activeness' (i.e. the number of messages being sent in a given period of time) allows for multiple topics to be brought up simultaneously, encouraging audience members to compete for floor control within the stream environment.

Finally, perhaps the most important factor is the live stream's content. Live streaming on platforms such as YouTube and Twitch often involves a broadcaster taking part in some activity, such as playing a video game, cooking, drawing, or promoting a product before their audience, while providing commentary and engaging with participants in the chat. However, to better examine the cross-modal strategies employed by streamers and audience members, I choose to utilize streams that focus on chatting and conversation between the broadcaster and the audience. In the past few years, streams in which the broadcaster is not live streaming a certain activity, but rather in which the focus of the broadcast is to interact and communicate with viewers, have grown in popularity. In fact, in March 2023, 'Just Chatting' is the most popular category on Twitch, with 14.9% of all hours viewed on the website falling into this category, beating out all other video games and activities. It seems clear that many broadcasters and audience members view live streaming not just as an arena to watch a streamer perform an activity, but also a venue for social (or in many cases, para-social) interaction between streamers and chatters. As I

aim to focus on the discursive methods prevalent in such interaction, the broadcasts used for this study can all be categorized as such.

However, as previously discussed, different platforms, channels, and broadcasts may have different expectations for formality and levels of engagement. I chose to break this down into formal and informal settings, and chose one stream that fell into each setting. Due to the aforementioned general expectation of informality on Twitch, I chose a YouTube broadcast for a relatively formal setting, but one where interaction amongst audience members is still encouraged. Future work may choose to look at even more strictly formal settings, such as an online university lecture or business webinar. However, such broadcasts often have privatized recordings or chats, and are less accessible to researchers. Therefore, the accessibility of public data remains a limitation of this study. It is possible in future work to further compare with such settings, especially in order to observe the influence of social power dynamics between participants.

For the purposes of this study, I selected two live streams for case study analysis, though my observations and conclusions will be contextualized with examples from a variety of other streams, though I have not formally performed a complete sampling of utterances within them. The two streams I selected for in-depth analysis differed along multiple axes: platform (YouTube Live vs. Twitch), size (‘small’ vs. ‘large’—such labels will be quantified further on), and social setting (‘formal’ vs. ‘informal’). This allowed for a broader look at the factors consistent across these axes, further supported by observations from other broadcasts.

3.2.2 Formatting case studies for CMDA

Once I had selected an adequate set of live streams to serve as case studies, I downloaded time-stamped transcriptions of the chat, as well as generated a time-stamped transcription of the audio. These transcriptions were generated using automatic speech recognition then corrected by hand. I then manually aligned the audio transcription with the time-stamped chat transcript to create a cross-modal

corpus that was ready to use for discourse analysis. Given that these are video broadcasts, these transcriptions notably lack the context of the video content. As Recktenwald (2017) points out, "in order to make sense of the interaction [between streamers and audience], it is necessary to construct some kind of multimodal transcription system that can account for what is going on in each of the socio-technical modes at any given point in time". For the case of "Just Chatting" broadcasts, we do not have key moments going on in the video channel that may impact topic generation and interactions, such as the streamer doing well or poorly in the video game or showing a piece of media to the audience. Nonetheless, a text-only transcription does fail to capture the video-based medium, in which acts such as body language and facial expressions often contribute to discourse. Thus, in addition to generating a transcript of the stream audio, I also watched the streams fully and made note of important physical actions on the transcript. Regardless, this limitation is still likely to affect the ability of this study to fully capture the entire range of discourse behaviors that may have been exhibited in these case studies.

Once the parallel corpora are generated, it is possible to perform the kind of qualitative analysis Herring describes by defining a set of phenomena and procedures and looking for examples of how they are carried out. It is also possible to perform quantitative analysis by performing a tag and count of each utterance as it relates to expected phenomena, but given the scope and time limitations of this study, a qualitative analysis is likely to yield more in-depth observations as well as provide the context that may allow us to reason about the interplay between technological, social, and discourse features.

Within each case study, I specifically chose to observe:

1) **Cross-modal interaction:** Any kind of dialogic conversation that occurs between the streamer and audience member(s) must by nature of the technological medium occur across the spoken and written channel. I observe these interactions, both in the written to speech form and the speech to written form, with respect to the strategies taken to facilitate these interactions, the timing of these interactions, and cases where these interactions fail or are incomplete.
2) **Parallel floors and independent topic generation:** I observe cases where chat members converse with one another rather than merely making monological speech moves or conversing

with the streamer, with respect to relevance and prioritization, disrupted adjacency, and independent topic generation.

3) **Topic management, development, and decay:** As with quasi-synchronous CMC, a high rate of topic generation and decay is hypothesized for the majority of live stream environments. Although it is difficult to quantify the rate of topic development, I tracked the decay of topics across both the spoken and written chat, as well as the abundance of irrelevant or weakly relevant contributions, to observe if the different channels behave differently in this regard. I also observed strategies related to how topics are changed (or how participants fail to change topics).

Quantitative observation can be carried out by creating an annotated corpus that documents cases of these phenomena and the types of behavior present, specifically how they align or do not align with expectations from text-based CMC and video-mediated CMC. The next section will cover findings from each case study and describe the phenomena of interest as they occur.

Although both streams were streamed publicly with the intention of drawing as many viewers from across the internet as possible, not all broadcasts are saved after the time at which they are broadcast. For example, Twitch streams can only be viewed for two weeks after their broadcast date by default, and become unavailable thereafter. This thus makes the collection of data from live streams different from the collection of data from a less ephemeral social media form, such as Twitter or blog posts, as although the broadcasts are publicly performed, in the case of Twitch live streams, they are not publicly stored. On the other hand, YouTube live streams do have publicly saved recordings that remain available on the broadcaster's YouTube channel. Therefore, while a Twitch broadcaster is undertaking a public performance, some may argue that not all members of the audience are doing so when sending messages, nor are they necessarily anticipating that their chat messages will be permanently recorded. Therefore, I will in the interest of user privacy anonymize all chat transcripts collected from Twitch broadcasts whose recordings are not uploaded for public viewing; because YouTube broadcast recordings are made public and permanent by default, I have not done the same for YouTube live streams. Given that

the broadcasters utilize live streams to perform micro-celebrity, and that removing all relevant information about them from the transcripts would be detrimental to the type of linguistic analysis I perform in ways that anonymizing audience members would not, I will not be anonymizing the broadcasters themselves in addition to audience members. However, I will remove active links or other easily identifiable information.

IV. CASE STUDIES

In this section, I will cover the set of utterances sampled from two case studies of live streams broadcast in 2020 and 2022 on YouTube and Twitch respectively. These cases differ along multiple additional axes: formality of setting, size and demographic of audience, platform, duration, and content. Given this, these case studies are not meant to be contrastive or argue that one set of phenomena is more likely to occur in one setting than another, as we cannot isolate one specific factor as responsible for a difference between the two case studies. Instead, this section focuses on similarities between these environments and suggests that online communities generally behave in such a way given the affordances of the specific platforms and technological features available to them. This section is organized broadly by phenomenon and discusses the qualitative observations I have made regarding my research questions for each specific case. Before diving into the specific phenomena and evidence from each case study, I will first generally describe the features of each individual case study and why these features are useful for this study.

Case study 1: Informal setting (relaxing and chatting)

This broadcast, streamed on Twitch in October 2022, is a ‘Just Chatting’ live stream hosted by an American entertainment streamer with approximately 20 thousand followers at the time. The broadcast host is a full-time streamer, meaning that her primary career and source of income is live streaming, and thus she must be well-versed in the typical etiquette and social setting of live streaming for financial as well as social reasons. She identifies as a female Dominican New Yorker who frequently uses both English and Spanish while interacting with users. The channel’s streams typically have a few hundred live viewers and a highly active chat, consisting both of frequent viewers and new viewers. I sampled a 100-minute section of this live stream, in which 1,964 chat messages were sent, for an average of nearly 20 chats per minute.

Twitch allows users to unlock special features, both channel-specific and global across Twitch, through paid subscriptions. Users may subscribe to channels by making a monthly payment to Twitch, with some of the money sent to the specific channel to which they are subscribed, to unlock channel-specific benefits. Users may also purchase general benefits from Twitch, such as a Twitch Turbo subscription, which allows users to watch live streams ad-free. Since Twitch displays whether each user in the chat is the streamer, a moderator, or a subscriber, I was able to get a sense of the roles of the most active participants. The majority of chat messages were sent by subscribers or Turbo users, comprising 1,382 messages, or 70.3%. 320 messages, or 16.2%, were sent by channel moderators. The remaining 262 messages (13.3%) were sent by users who were neither subscribed nor affiliated with the channel. Note that such users may still be active followers of the channel; not being subscribed simply means they do not make payments to this channel in particular. Twitch also allows certain channels to upload and use custom emojis; while default emojis are rendered in emoji form in my transcript, custom emojis will be rendered only as their names. Another feature of Twitch chat is the ability to reply directly to another user's message, which we will find is highly relevant to our discussion of topic generation and relevance. Such messages are rendered in this transcript with the username of the user being replied to at the beginning of the message. In a live Twitch chat, these messages would be rendered with an indicator to the message to which they respond.

The below excerpt from the chat transcript exemplifies the usage of some of the aforementioned features. In message 1, a user uses a bot command (`!lurk`) to indicate that he will no longer be active in the conversation, followed by an explanation of why and several custom channel emojis, and in message 2, a moderation bot sends an automated reply. In message 5, a default Twitch message is sent indicating that a user has purchased a subscription to the channel. In message 7, another user uses the reply function to respond to message 1.

As mentioned in the methodology section, this example and all further transcripts from this case study have been anonymized. Twitch users' usernames will be replaced with generic letters that are reused across examples; i.e., unless otherwise stated, `USER A` in a given example is not necessarily the

same individual as `USER A` in another example. Moderators of the channel will be denoted with the word `(Moderator)` preceding their messages.

Example 1

#	Timestamp	Chat
1	0:45:53	<USER A> !lurk Mannnn. It's aight Im headed out anyway. Gotta 4 hour CE lecture to attend to qnovaWave qnovaWave qnovaWave qnovaWave
2	0:45:53	<BOT> that lurk is greatly appreciated, you go on and have a great ass day!
3	0:45:55	<USER B> qnovaNoted
4	0:46:00	<USER B> I knew it!
5	0:46:03	<USER C> subscribed with Prime. They've subscribed for 5 months! <3
6	0:46:08	<USER D> qnovaLookBack qnovaUwU
7	0:46:10	<USER E> @<USER A> have fun

In this broadcast, the live streamer interacts with the audience through the video stream, which consists of a webcam capture of herself. Unlike the majority of live stream categories on Twitch, in which a host or performer is undertaking some activity (e.g. gaming, cooking, drawing, etc.) for the audience to watch, the primary goal of a Just Chatting live stream is generally to interact with the audience or carry on a conversation, without an additional layer of activity. This may make these broadcasts less complex with respect to the number of ongoing conversations and interleaved topics and meta-topics, but at the same time makes them more likely to have active conversations going on between the broadcaster and audience, both features that are beneficial for the current study. Additionally, the stream may not have an explicit goal or topic, and instead is mainly purposed for general conversation. This may influence the rate of topic development, similar to a general instant messaging chat room or forum thread.

Case study 2: Formal setting (educational/tutorial)

The second case study I selected is a YouTube Live broadcast, which streamed live in April 2020. The broadcast host is an IT and cybersecurity career development and tutorial YouTube channel with over 250,000 subscribers at the time of writing. However, the live broadcast only had 1,800 views and a total of 275 chat messages over the course of its 1 hour and 13 minute live time. Of them, 46 are sent from channel moderators, who fulfill a similar role to channel moderators on Twitch. Unlike the Twitch stream, the recording of this broadcast was uploaded to the channel afterward, which by default happens automatically to YouTube streams.

This stream, entitled "AMA while I setup some Virtual Machines", takes on a question-and-answer format signified by its title (AMA standing for "ask me anything"). Similarly to the first case study, there is no designated topic of discussion, but the channel's general audience and the authority upon which the broadcast host has built their brand creates a social impetus to keep all questions and contributions 'relevant', despite the lack of explicit rules defining the range of acceptable topics.

Unlike the first case study, in which the main activity the stream is centered around is conversation, this live stream involves the broadcasting of a particular activity: the broadcaster demonstrating how to set up a virtual machine on his computer, while simultaneously conversing with the audience about general topics related to IT and computer systems. Therefore, the actions taken by the broadcaster, which are streamed to the audience, necessarily impact—and frequently interrupt—the timing of cross-modal interaction with the audience. Although the text transcript does not account for it, I will discuss such actions where relevant, in order to place the conversational behavior in context with the ongoing activity of the live stream.

Like the first case study, this live stream is characterized by a question-and-answer form of interaction, in which the broadcaster similarly manages multiple ongoing conversations with different viewers through exchange chaining and the explicit signposting of conversational topics. However, the existence of a 'baseline' activity (demonstrating the setup of the virtual machine) to which the broadcaster

returns between such exchanges creates another axis of relevance both for the broadcaster and the audience to mentally keep track of. This ongoing activity can be viewed as yet another conversational floor that involves the speaker as a presenter and the chat as an interactive audience to the demonstration and description of this activity, which takes place both in the spoken and written channel.

Greetings and mutual orientation

In most forms of voice- or video-based CMC, mutual orientation is established before conversation begins. In the case of live streams, users frequently enter and exit the chat space, often without necessarily sending an explicit greeting or goodbye message. As a result, the task of greeting and establishing orientation within the space falls almost exclusively onto the broadcast host.

In case study 1, one of the first prominent occurrences in this broadcast is greeting and the establishment of orientation within the broadcast environment. Unlike text-only CMC, which is often observed to omit the need for greetings, greetings appear to be quite standard in the live stream environment, and occur not only at the beginning of the stream, but also throughout as additional audience members join the conversation. Example 2, an excerpt from the very beginning of the stream (the first ten minutes constituted a 'waiting room' period, in which audience members may send chats, but the video broadcast has yet to officially begin) demonstrates interaction between the broadcaster and the chat wherein the broadcaster addresses the chat both as a collective and individually greets select individuals.

Example 2

Timestamp	**Video Transcription**	**Chat**
0:09:41	Hey y'all hello	<USER A> sup chat qnovaWave qnovaWave qnovaWave qnovaWave
0:09:45	How you doing?	<USER B> @<USER C> qnovaWave
0:09:48	How y'all doing today? I hope well	<USER D> !drkskn
		<USER E> qnovaWave qnovaWave qnovaWave qnovaWave qnovaWave

		<USER F> Hiiiiiii
0:09:51	I hope y'all doing well today	<USER D> HI NOVA!
0:09:52	Talk to me, chat	<USER F> qnovaCute
0:09:54	Say something	<USER C> Hey Nova! qnovaWave qnovaWave qnovaWave
0:09:55		<USER G> herro
0:09:56	There we go [inaudible]	<USER H> !yencid
		<USER I> qnovaWave qnovaWave qnovaWave hi strimer
0:10:04	Hi <USER G> hi <USER C> Hi <USER I>	<USER E> Yoooooo qnovaWave qnovaWave qnovaWave qnovaWave qnovaWave qnovaWave
0:10:04	[music clip plays]	<USER J> novvvva qnovaWave qnovaWave qnovaWave qnovaWave qnovaWave qnovaWave qnovaWave qnovaWave qnovaWave qnovaWave qnovaWave
0:10:05		<USER K> whats poppinnnnnn
		<USER C> @<USER D> qnovaWave qnovaWave
		<USER A> qnovaWave qnovaWave qnovaWave qnovaWave qnovaWave
0:10:06	You just – you love to ruin my mood, huh	<USER L> Hey Nova yerrrrrrrrrrrrrrrrrr
0:10:10	You know I come up in here excited as fuck	<USER M> qnovaWave qnovaWave qnovaWave qnovaWave
0:10:11	And you play that?	<USER F> Lmaooo
0:10:12	Are you kidding me Don't do that to me	<USER H> I LOVE YOOOUU TOOOOO
0:10:15		<USER C> @<USER K> qnovaWave
		<USER H> HI NOVA

The same form of greetings occur at the beginning of case study 2, though due to the smaller size of the audience, far fewer greetings occur. The conventions for what exactly chat users say at the beginning of the live stream differs between the two cases; while in case study 1, the Twitch audience uses emojis and phatic greeting messages to greet the streamer and the rest of the chat, the YouTube audience comments "`first`", "`second`", and "`Thoid` [sic]", a common practice done by users who are among the first to view a certain YouTube video or broadcast. Nonetheless, the greeting and response exchange is very similar between both cases, with the live streamer individually greeting certain audience

members while also addressing the audience as a whole with address terms like “y’all” or “chat” (a common term used by Twitch streamers to refer to the audience as a collective).

Example 3

Timestamp	**Video Transcription**	**Chat**
0:00	Anyway hope you guys are doing alright just put out a video well like what	<(Moderator) KevRunsOnDunkin> first
0:06	ten minutes ago hopefully you guys came in for the live chat if you didn't at	
0:13	the end I said “I think I'm gonna do a live stream” and here we are live streaming	<Proactive Progression> second
0:19	so that's exciting I'm trying to view this thing and it	
0:29	ain't workin’ “view on YouTube” just want to make sure this thing’s working properly here ah, there we go	<Johnny CincoCero> Thoid
		<LVC> yo
0:40	What’s up y’all? What’s up KevRunsOnDunkin, [inaudible], Johnny, LittleVikingCoach	
0:47	How y’all doing today? If I can find the right screen, just in this livestream, we’re gonna do an ask me anything	<Elyziah Reyes> Live stream of what

As observed in Recktenwald (2018), the majority of speech acts in live streams are monological or constitute incomplete exchanges. This is likely due to the inability of the one host to respond to all conversational exchanges initiated by the audience. Moreover, not every audience member will choose to respond to a given utterance by the streamer, such as “How y’all doing today?”

Greetings do not only occur at the start of the stream. In case study 1, as the stream goes on, the streamer continues to greet frequent audience members or those she recognizes when they send messages. Sometime this is in response to a greeting message sent by the audience member, as in example 4 below, but much more frequently, audience members simply jump into the ongoing conversation using the text channel, as in example 5.

Example 4

0:54:31	So on Fanhouse, I was like, for every like I will buy a drink,	<USER A> we watching at work
0:54:34	and there's 10, and I'm like, I'm not buying 10 drinks I can't do this	**<USER B> Heyyyy wasss popinnnn**
0:54:40	**<USER B> what's good? Your name looks super familiar**	<BOT> Sneak peeks at merch/emotes, videos of my life, selfies, photography, memes, group chats HERE - sub to my FanHouse <PROMOTIONAL LINK>
		<(Moderator) USER C> so you lied?
0:54:42	**Welcome in**	<USER D> SCAMMER

Example 5

0:52:42	Modern Warfare 2 bundle on PlayStation direct	<USER A> you just gotta sign in to buy and its $560 before taxes
		<USER B> GOW bundle?
0:52:45	Right now? Oh then I might have to	<USER C> MW2? qnovaTomato
0:52:49	Uh. Modern Warfare 2 Tomato? OK **Hi <USER B> hello** You gotta sign in to buy and it's 560 before taxes - oh	<(Moderator) USER D> @<USER E> alrighty qnovaDone

It appears from both case studies that greetings are highly common in live streams and frequently constitute largely complete, albeit interrupted, interactional pairs between the broadcaster and individual chat members. At the same time, greetings frequently exhibited a conversation between the broadcaster as

an individual and the audience as a collective. In the larger audience of case study 1, my findings echo on a smaller scale findings from past work on mass Twitch audiences, such as Ford et al. (2017), who suggest that the collective 'crowdspeak' formed by large audiences of live streams utilizing short messages with little unique lexical content manages to produce coherence as a collective. Such a phenomenon is usually observed in massive Twitch channels with viewers in the tens of thousands, and usually occurs when the audience is collectively viewing an activity (i.e. in a video game stream, for example, rather than a chatting stream). However, the chat samples from my case studies, in particular case study 1, suggest that a collective back-and-forth between the streamer and the audience can occur when it comes to common conversational practices, such as greetings.

Cross-modal interaction

The primary form of conversation in the broadcast centers on the interactions between the streamer and the audience, either as a collective or with individual audience members. Utterances made by the streamer spawn multiple reactions in the text chat at once, as is expected of the one-to-many structure of the conversation. However, the existence of multiple parallel channels that occupy different mediums means that conversation can continue in the text chat parallel to a different topic in the spoken channel, meaning that multiple branches of conversation can be simultaneously entertained in the text channel. This differs from face-to-face conversation, where this type of branching cannot be sustained. For example, the streamer might raise a question to the audience in general, where a range of responses is anticipated and wanted. Unlike a face-to-face presentation, however, not only can multiple audience members respond simultaneously, the presenter can also entertain multiple possible responses simultaneously, due to the nature of written text to maintain a record of all past responses. The below example shows the presenter posing a request to the audience as a collective for advice, and follows by responding to individual audience members.

Example 5

0:14:55	Help me pick a chocolate	<USER A> **All of em**
0:14:56		<USER B> The inner NY is coming out I see and what did I do Cheer100
0:14:58		<USER C> **Eat the red candy first**
0:14:59		<USER D> dnelltLaughing
0:14:59	Don’t give me a question mark <USER E> don’t do that	<USER E> ??
0:15:00	Help me pick a chocolate	<USER F> ‘TheseDudes’ qnovaBoujeeSip
0:15:02	What chocolate	<USER G> **doesn't matter which chocolate they're all good**
0:15:04	**All of them? I can't**	<USER H> **the red 1**
0:15:05	<USER> thank you very much for the 100 bits	<USER I> **red**
0:15:06		<USER J> qnovaWave qnovaWave qnovaWave qnovaCute qnovaCute
0:15:07		<USER K> **Red**
0:15:09	**Doesn't matter which chocolate they're all good**	<USER L> **Yo nova eat the red one**
0:15:14	Hi <USER F> hi <USER D>	<(Moderator) USER M> the square ones @<STREAMER>
0:15:14		<USER N> no dap is crazy🤨
0:15:16	**Eat the red one?**	<USER O> i’m allergic so none
0:15:17		<(Moderator) USER P> cute is an understatment. give yourself more credit superstar qnovaCute
0:15:18	What's the red one let me see	<USER E> B4
0:15:21		<darthbeeta> Did my granny knit that hahha okay im done
0:15:22		<USER A> You might not want to pick them all, but you can. And thats what matters.
0:15:24	Let me see what the red one is	<USER I> Actually White
0:15:27	What is that one	<USER Q> red one yeh
0:15:27		<USER D> qnovaWave
0:15:29		<USER R> Hiiii novaaa qnovaWave krystaaHi qnovaNom
0:15:32	[reading] Hazelnut and chocolate paste covered with cocoa powder	<%USER F> **How about you just eat the 2nd one in the first row**
0:15:40	OK I'll eat that one sure	<USER S> white looks good

0:15:41	**How about you just eat the 2nd one?**	<USER T> there's a book to tell you which chocolate there is?
0:15:44	**Which 2nd one**	

In the above example, cross-modal interaction is depicted through the color-coding of specific interactions. In the literature on CMC, the interleaving nature of text-based online communication is frequently depicted in a variety of manners, including arrow-based diagrams and physically separating the transcript into multiple columns representing parallel conversations. For my purposes, I will maintain a two-column format to clearly distinguish between the broadcaster's speech and audience messages, and instead use color to represent distinct threads of conversation. When the broadcaster responds to a specific audience member's comment or vice versa, I highlight the interaction in the transcript in the same color to demonstrate that, despite the lack of adjacency, they constitute one interactional pair.

Example 5 from case study 1 indicates how the broadcaster generally manages cross-modal interaction. Given that many messages are sent simultaneously in the chat, the live streamer must respond to them in a manner that organizes her responses such that it is either apparent to what message she is responding or otherwise ensuring that her response is taken appropriately regardless of to what or whom she is responding. Given that the spoken channel constitutes the 'main stage' of the live stream discourse environment, and the written channel the 'backstage', it follows that a given utterance made by the streamer could equally be likely to be a response to any recent chat message, and thus must be explicitly signaled; meanwhile, chat responses do not need to explicitly call out the exact utterance to which they are responding, as optimal relevance is assumed to the ongoing topic (or global set of topics) being discussed in the spoken floor.

As the above excerpt shows, a few different strategies are used to index the relevant prior message from the chat and turn it into the topic on the spoken floor, a practice that Recktenwald (2018) refers to as 'topicalizing'. In the following examples, I will show samples that exemplify these strategies. When each example depicts only one chat participant sending multiple messages, I will anonymize these participants simply as <USER>.

1) addressing the speaker of the prior message by name:

Example 6

0:14:49		<USER> ?
	.	
	.	
	.	
		.
		.
	Don’t give me a question mark	.
0:14:59	<USER> don’t do that	

Example 7

0:56:10		<USER> WOAH YOUR HEADSHOTS NOVAAA HALOOO <3
	.	.
	.	.
	.	.
0:56:19	Ah thank you <USER> thank you very much	
0:56:20	I appreciate it thank you thank you	

2) directly quoting the prior message:

Example 8

0:15:29	.	<USER> How about you just eat the 2nd one in the first row
	.	
	.	
		.
	How about you just eat the	.
0:15:41	2nd one?	.
0:15:44	Which second one	

3) echoing the prior message while index shifting:

Example 9

0:56:19	. . .	<USER> Are you excited to finally play on PlayStation 5?
		. . .
0:56:33	Am I excited to finally play on PlayStation?	
0:56:34	I mean not really.	

The act of shifting the index pronouns, frequently from the second person to the first person, as shown in the example 9, occurs extremely frequently, and especially when the live streamer intends to immediately respond to a question regarding a personal attribute, goal, or intention. This behavior is consistent with past discourse studies of other live streaming platforms, such as Licoppe & Morel (2018)'s work on the live streaming platform Periscope.

The expectation of optimal relevance to the spoken channel can also be seen in the different expectations for acceptable delay when it comes to speech to written interactions vs. written to speech interactions. Speech to written interactions are observed to have up to nearly 3 minutes of acceptable delay; the relaxed expectations for sequential relevance in the text channel allow for text responses to spoken utterances to occur with a high amount of delay and often with very minimal relevance.

However, this can result in ambiguity when the spoken floor has changed topic but the written channel has not. For instance, in the below example, the 2-minute delay causes ambiguous anaphora resolution. An audience member refers to a specific video game franchise that will be receiving a new release, and refers to the same game again in another chat utterance two minutes and three seconds later. Though such an utterance might be acceptable in a one-on-one conversation, the one-to-many structure of the live stream requires the streamer to attend to other topics of conversation with other audience members during this time. Therefore, while the anaphoric antecedent is salient to the specific audience

member making this remark, it is unlikely to be for the broadcaster or other members of the current conversation, resulting in ambiguity:

Example 10

	.	
	.	
0:53:28	.	
		<USER> Crisis Core is coming out this year
		.
		.
	Crisis Core is coming out this year?	.
0:53:33	When is it coming out	
	.	
	.	
0:55:31	.	<USER> and you SHOULDN'T play it unless you want to be majorly SPOILED!!!!
	But yeah. And you shouldn't play it unless you want to be majorly	
0:55:42	spoiled?	
0:55:45	Wait, what?	
	.	.
	.	.
	.	.
0:56:12		<USER> Crisis Core you shouldn't play it

Moreover, the above example exemplifies the high percentage of incomplete discourse exchanges; due to the number of ongoing conversations that must be managed primarily by one individual (the broadcast host), it is likely that many exchanges will be completely dropped. Though the audience member continues to carry on the conversation in hopes of receiving a response, the broadcast host does not return to this exchange in favor of carrying on other exchanges with other audience members or the audience as a collective.

On the other hand, the transience of the spoken stream (as opposed to the record-leaving text chat) as well as the ‘main stage’ nature of the spoken floor impose a higher restriction on the temporal

relevance of the broadcaster's responses. While the broadcaster may bring up an old topic, they almost exclusively do so by responding to a new message. If a message is not responded to immediately, it likely will not be at all, and generally, no greater delay is observed between a text message and an indexed response from the broadcaster than a few seconds. Though there may simply be a lack of evidence to the contrary, further evidence lies in the frequency with which users make use of self-repetition as a strategy to gain notice and thus topic control of the spoken stream temporarily, if the first iteration of the message fails to gain notice. The following series of messages show one chat participant at separate time intervals not only engaging in self-repetition, but also indicating a past message that was ignored by the broadcaster.

Example 11

0:49:38	. . .	<USER> @<STREAMER> She wanted to see if we can chill one day and I was like for sure. We. exchanged numbers and I left. I texted her the next day but since she a terrible texter she hasnt replied to my last message i sent so im wondering if i should hit her up straight up and make plans. Thoughts?
0:51:52	. . .	. . . <USER> posted the end of the story fyi
	. . .	. . .
0:54:42		<USER> @<STREAMER> She wanted to see if we can chill one day and I was like for sure. We. exchanged numbers and I left. I texted her the next day but since she a terrible texter she hasnt replied to my last message i sent so im wondering if i should

		hit her up straight up and make plans. Thoughts?

Contrast the highly synchronous nature of written-to-speech interaction with that of written-to-written interaction. This further adds to the main stage vs. backstage distinction between the two modes as well as the characterization of cross-modal interaction as being the primary goal and method by which interaction is carried out. Interaction and conversation between viewers is considered secondary, although still an important aspect of the live stream environment, and will be discussed further in the following sections.

Topic development & decay

Given that quasi-synchronous text-based CMC is observed to sustain multiple interleaving topics, but spoken conversation is not, it is expected that the structure of live streams is likely to be similar to the branching structure observed of video-mediated communication. An utterance in the spoken channel might set off a variety of responses in the text channel, but not every response can be addressed in the spoken channel. This facilitates the generation of the text channel as a parallel floor where independent conversations can exist as the 'primary' topic of conversation goes on in the spoken channel.

However, what was actually observed in the vast majority of this stream is that multiple interleaving topics are sustained across both channels, with the streamer moderating the shift between topics by responding intermittently to text-channel messages on any one of the topics. As studies of quasi-synchronous CMC would lead us to expect, a given utterance by the streamer results in a wide range of responses in the text chat, leading to a range of possible next topics. However, the streamer responds by herself modulating between the set of salient topics and responding to relevant messages as they are sent, resulting in a similar interleaving of topics even in the video channel. This implies that the nature of quasi-synchronous CMC to sustain multiple threads of conversation either is not unique to the

written nature of the medium, or is capable of impacting other parallel conversations that occur alongside the written channel.

Below is an example from case study 1, where at 0:49:20, the stream host presents a new topic: asking the chat for advice on purchasing a PlayStation 5 console.

Example 12

0:50:42	So should I	<USER A> boy
0:50:43		<USER B> god of war comes out when nova says
0:50:45	Should I buy the PlayStation 5 as is at a wild 700 dollars right	<USER C> Spanish novaa is more shy and humble
0:50:55	'cause I don't mind	<USER D> Are you getting it on PlayStation Direct?
0:50:56	I'll do that shit	<(Moderator) USER E> @<USER C> qnovaWave qnovaWave qnovaWave
0:51:01	Or should I wait to be invited by Amazon, 'cause they do invitations and I can get one like at a regular price	<USER A> **700?**
0:51:06	So yeah	<USER E> qnovaNoted
		<USER F> @<USER C> qnovaWave Hi
0:51:08	**Like seven - yeah it's like 700**	<USER G> Oooo
0:51:09	**dollars to get a PlayStation 5**	<USER G> Maybe
0:51:09	But it's okay because it was	<USER H> wait
		<USER I> you should wait lowkey
0:51:12	funded so like I'm willing	<USER J> Please try to get into Playstation Direct
0:51:13	to do that	<USER K> Regular for sure

The conversation about PlayStation games goes on until around 0:52:03, at which point the streamer switches to continuing a past conversation with a particular viewer in Spanish. After responding in Spanish, she immediately returns to the conversation about games and consoles, until the conversation is interrupted again by a response to an ongoing conversation with a different viewer and a greeting of a new viewer. This fragmented format of speech is carried on by the streamer without explicit signaling of conversational shifts. She may indicate the intention to respond to a specific message using the methods I

identified in the previous section, but does not need to signpost a change in conversational topic. As a result, the spoken channel shows a similar lack of sequential coherence as is frequently observed in text-only CMC.

Example 13

0:53:00	I shall do that	<USER A> **nova live qnovaGasp**
0:53:01	Uh thank you for the follows everybody	<USER B> better wait till them black friday deals
0:53:04	Hello hello welcome in. **Nova live - Yes, I am**	<(Moderator) USER C> see i told you these nerds would know
0:53:09	**Hi <USER A>.** **<USER D>, I asked you a question**	<USER D> @<USER E> <USER E> senpaiiHEARTBEAT senpaiiHEARTBEAT
0:53:10	and I'm so sorry I probably didn't see it, let me see	<USER F> Final Fantasy 7 rebirth is not coming out this year it's next
		<USER G> **You're on super early Nova! Yeah Playstation 5 are available on Amazon, Walmart, PS Direct, and theough Gamestop as bundles. The only thing they're doing special for God Of War 2 is a special controller!**
0:53:13	Let me see Miss <USER D>	<(Moderator) USER H> the nerds LMAO
0:53:15	**Wait <USER D> yeah do you speak**	<USER I> lisabbFine xxboxx ?
0:53:16	**any other languages? Other than English?**	<USER D> **OMG I MISSED IT**
0:53:21	**You missed it? It's OK**	<@(Moderator) USER C> @<(Moderator) USER H> qnovaEvil
0:53:22	**Hi Locker hello**	<USER D> **SI !**
0:53:24	**Yeah PlayStation5s are on Walmart, Amazon and PlayStation direct**	<USER G> You right, I am a proud nerd @<(Moderator) USER C>
0:53:27	**And through GameStop as bundles. Only thing they're doing special**	<USER E> **she speaks latin**
0:53:28	**for God of War 2 is a special controller. Ahh**	<USER F> Crisis Core is coming out this year
0:53:30	**OK OK**	
0:53:31	**Si ah so you speak Spanish as well**	

As shown in the above example, the broadcaster easily interleaves multiple topics of discussion, responding to new chat messages as they come in, and returning to past topics when relevant. Therefore, the conversation takes on an interleaved structure, where new utterances are relevant to past ones with a few degrees of delay, resulting in the interleaving of conversational topics and a lack of sequential coherence.

Something to note in the above excerpt is the streamer's engagement of conversational repair with regards to an ongoing thread of conversation she is hosting with a particular viewer:

Example 14

0:53:09	<USER D>, I asked you a question	. . .
0:53:10	and I'm so sorry I probably didn't see it, let me see	
0:53:13	Let me see Miss <USER D>	
0:53:15	Wait <USER D>yeah do you speak	
0:53:16	any other languages? Other than English?	<USER D> OMG I MISSED IT
0:53:21	You missed it? It's OK	. . .

The streamer points out a missed message while simultaneously keeping up another track of conversation, both with the audience member in question and with the rest of the audience.

As this exemplifies, the streamer replies to chat messages as they come in chronologically, which means that the interleaving nature of topics in the chat carries over to the video stream. This differs from what observations of video-mediated communication might lead us to expect, where the spoken channel sustains one thread of conversation (that is generally considered the 'main' topic or at least the most salient one) while many branching threads can be entertained in the 'backstage' text channel. What we see here instead is that the spoken channel simultaneously engages in multiple threads of conversation. We

might hypothesize that the social structure of the live stream disincentivizes independent conversations in the text chat, as conversations are expected to be oriented towards the live stream and interaction with the broadcaster, especially in a 'Just Chatting' form of live stream, in which the broadcaster is viewed less as a performer and more as an individual for the audience to intimately interact with. At the same time, the technological structure of the live stream may have an impact: unlike the 'democratic' structure of a video chat, where any user who has microphone access may hold the spoken floor, the broadcast host has a monopoly over the spoken floor in a live stream. As such, there can be no situation where another user or topic takes over the spoken floor and the previous speaker must make the rest of their contributions in the written channel. The broadcaster has complete control over the spoken floor, and as such, dictates the direction of conversation. In this case, the broadcaster does so both by independently raising topics of conversation and by selecting topics proposed by the audience and bringing them to the main stage of the live stream.

Interaction and topic development follows a similar pattern in case study 2. The designated "Ask Me Anything" format of the live stream formalizes this concept, with the broadcaster engaging in one-on-one exchanges with audience members, many of which are ultimately incomplete, with either the streamer or audience member dropping the exchange after rarely more than one or two follow-ups. Therefore, despite the seemingly more formal nature of the setting, topic decay occurs at a similarly high rate, with the broadcaster responding to new questions as they come in. Past topics of conversation occasionally receive follow-ups, which, like in case study 1, may be interleaved with newer topics according to the broadcaster's choice of moderation. However, they will eventually fail to be sustained as the number of new topics grows, resulting in fast rates of topic decay and sequentially interrupted relevancy. The below excerpt exemplifies this question-and-answer style exchange chaining, with the initial exchange (with user `Erel H L`) being incomplete, as the audience member fails to respond to the question.

Example 15

Timestamp	Video Transcription	Chat
0:5:37		<Erel H L> what are pros & cons of MSP or in-house department. which do you recommend?
0:5:46	yeah it'd be probably much easier on Intel What are the pros and cons of MSP	<Ali Smith> Hi zach , i really appreciate make a stream at time staying home situation
0:5:56	or in-house department, which do I recommend? Um as far as what though? As far	<x0Sudo> as A+ worth in 2020 ? <(Moderator) KevRunsOnDunkin> But ram as zach says
0:6:02	as what pros and cons	<No Degree I.T. !> Its a pain for sure to get mac os on a virtual box <derekv81ify> i'm looking to switch careers from a factory assembly worker to IT, can you point me in the right direction? assuume i have 0 experience and just give whatever advice you think could help
0:6:05	Hi Zach I really praise you mate really appreciate make a stream at time staying home situation For sure brah	
0:6:10	Is A+ worth it in 2020? Yeah I think the A+ is gonna be worth it for a very long time	<(Moderator) Megladon> @Cadence Rudd yes but you dont need vm for that <LVC> you said no bootcamps,better options?
0:6:20	Uh Derek you're looking to switch careers from factory assembly worker to IT can you point me in the right direction	
0:6:26	Assume you have zero experience and just give whatever advice I think could help Well quite honestly I've done - I mean	
0:6:33	probably at this point in time like at least 50 videos talking about this which is why I created this course	

The rate of topic development and decay in the much more formal YouTube Live stream suggests that the informal and relaxed setting of case study 1 was not the determining factor in creating the sequentially disjointed nature of conversation I observed. Moreover, case study 2 featured a much less fast-paced chat with fewer participants, suggesting that it is not the case that disrupted exchanges occur only when the size of the broadcast necessitates the dropping of exchanges by the broadcaster. Instead, the rate of topic decay may have much more to do with 1) the inherent one-to-many nature of conversation and 2) the delay and interruption to traditional turn-taking caused by interacting across a synchronous, instant medium and a quasi-synchronous, delayed medium.

Parallel Floors and Independent Topic Generation

In studies of VMC, it is frequently observed that independent topics may appear in the text chat when conversational topics are branched off from the spoken channel or when speakers who previously conversed in the spoken channel move to the written channel to yield the spoken floor while carrying on their interaction in a backstage channel. Of course, such interaction is impossible in the live stream setting where audience members are not able to contribute to the spoken channel, and is moreover unlikely given the social structure of the live stream that is oriented around the streamer. Nonetheless, I observed some instances of audience members conversing with each other, but these conversations were usually short-lived, and usually had to be explicitly signaled with direct address or the reply function. Given that the chat was far more active in case study 1, I found several more examples of chat members interacting with each other, as shown below. In example 16, the reply function is used to specifically designate a user's past message through displaying a reply on Twitch. In example 17, no specific technological feature is used in the first message, but the second user, who is a moderator of the channel and thus a frequent member in the community, is referred to by name.

Example 16

```
[0:56:12] <USER A> Crisis Core you shouldn't play it
...
[0:56:56] <USER B> @<USER A> Crisis Core is the prequel to the main story
```

Example 17

```
[0:13:49] <USER A> damn <USER B> she said you aint got no drip
...
[0:14:21]  <(Moderator) USER B> @<USER A> everyone lies vylinShrug
```

Conversation amongst audience members seems more likely to be sustained if participants are familiar with each other or if there is a lull in interactive conversation in the stream, suggesting that though the multi-modal nature of live streams facilitates the generation of parallel floors, the presenter–audience dynamic centers the discursive environment around the streamer as a presenter, performer, or moderator. Sustained conversation in the text channel thus remains rare; even though new and independent topics are frequently introduced by audience members, they are often attempts to gain control of the spoken floor through notice from the streamer, and very rarely attempts to initiate new contributions to the written floor. For instance, the below chat messages are common examples of new topics initiated by audience members, who must make such contributions relevant to the streamer, even if they are independent topics unrelated to the current live stream, in an attempt to gain a response and thus momentary control of the spoken topic.

Example 18

```
[1:03:48]   <USER> Playing any games with the chat today???
```

Example 19

```
[0:38:55]   <USER> Nova I'm boutta make pizza, what should I put on it?
```

Such topics are thus unlikely to receive responses from fellow audience members if they are not brought into the stream through acknowledgment from the streamer. Independent conversation can thus only occur if 1) a question or issue related to a topic discussed on the stream (or the stream itself as in example 17) is raised or 2) if the topic of conversation remains consistent with the stream topic.

Example 20

```
[0:57:56]  <(Moderator) USER A> I cant even redeem channel points
           redeems????
...
[0:58:31]  <(Moderator) USER B> @<(Moderator) USER A> thats so strange wtf
...
[0:59:02]  <(Moderator) USER A> @<(Moderator) USER B> yea i gotta brb this
           is driving me nuts i have wild things to say and i cant say
           them the way i want
...
[0:59:36]  <(Moderator) USER B> @<(Moderator) USER A> LMAOOO alrightyy
...
[1:02:06]  <(Moderator) USER A> @<(Moderator) USER B> it aint working, i
           gotta fight twitch
...
[1:02:32]  <(Moderator) USER B> @<(Moderator) USER A> wtf thats so
           strange..... wut is going AWH
```

Example 21

```
[0:57:47]  <USER A> I wanna try and go out to NY next year so fingers
           crossed
...
[0:59:18]  <USER A> im just tryna spend time in the city and hit different
           spots, get food and drinks, see some shit lol. Nothing super
           specific.
...

[0:59:38]  <USER B> @<USER A> pulllupppppp
...
[1:00:07]  <USER A> @<USER B> I wanna try to hit a Nets game too qnova10
...
[1:00:20]  <USER C> @<USER B> based @<USER A>
...
[1:00:38]  <USER B> Netss????? You mean Knicks right @<USER A>
...
[1:01:14]  <USER B> It could be Knicks vs Nets how about that, in MSG
           @<USER A>
...
[1:01:22]  <USER A> @<USER B> LMAOOOO fasho
...
[1:02:47]  <USER A> If I go out to New York, whoever I meet / visit gotta
           come out here to San Diego in the summer and do the same
           imshigHearts
...
[1:03:10]  <(Moderator) USER D> @<USER A> in the summer? you're sadistic
...
[1:03:39]  <USER A> @<(Moderator) USER D> okay fall or winter then lmaooo
...
[1:04:14]  <(Moderator) USER D> @<USER A> fall is perfect cuz its still
           warm for like another month and a half
...
[1:04:33]  <USER A> @<(Moderator) USER D> that Cali fall is beautiful
```

Recktenwald (2018) writes about the nature of intra-audience communication in Twitch streams that “the intra-chat communication is very anonymous and there is very little phatic communication between participants”. This appears to be in contrast with the examples shown above, and further evidence from this case study, which shows users greeting and saying goodbye to one another, as well as responding to each other’s messages with emojis and phatic reactions.

Example 22

```
[0:40:21]  <USER A> Hi everyone alexotCozy alexotHype
...
[0:40:38]  <USER B> @<USER A> <USER A> qnovaLookBack qnovaHeart qnovaHeart
...
[0:40:41]  <USER C> @<USER A> qnovaWave qnovaWave
...
[0:40:45]  <(Moderator)> @<USER A> hiiiiiii <USER A> qnovaWave qnovaHeart
```

Such actions, however, appear to primarily occur amongst the most frequent chatters as well as channel moderators, suggesting that there is an additional level of in-group status conferred to those who are recognized as frequenting the community, and thus interact with each other more frequently and less anonymously. However, this may also be evidence to suggest that, as far as the evidence from this case study suggests, conversation in live stream settings is viewed as a group activity; although conversation is implicitly assumed to be centered around the topics discussed on the stream, and audience members expect to center their experience around the streamer, it is not the case that each audience member is independently expecting their experience to be a one-on-one with the streamer in which the other audience members only exist to compete for time and notice. Instead, intra-chat interaction is seen as a fundamental part of the social activity.

In case study 2, there does appear to exist some kind of established community within the audience, as certain audience members do seem to recognize and greet one another, though this occurs to a lesser extent than in case study 1. The much smaller size of this audience means that chat messages can be longer, can have longer expected delay, and can stay on the screen longer, allowing users to have a

better chance to interact with one another. However, the more formal social setting and question-and-answer designation of this live stream may inhibit the natural generation of conversation between users. Given that users are encouraged to ask questions and advice of the broadcast host, whose brand and audience rests upon the acceptance of his authority as an IT career coach, participants answering questions or giving advice to other participants may be viewed as a face threatening act. Thus, the only interaction that occurs between audience members can be social, which is only observed between audience members who appear to be familiar with one another.

Example 23

```
[16:37]  <No Degree I.T. !> another legend in the chat !!!!
[16:41]  <No Degree I.T. !> whats up du'an
[17:03]  <(Moderator) Du'An Lightfoot> What's up No Degree.
[18:12]  <(Moderator) Du'An Lightfoot> Thanks for all you do bro!
[18:42]  <No Degree I.T. !> Du'an i finally found a use for my pi4 taught
           my son how to set it up so he could do school work on it while
           stuck at home!!
           . . .
[19:41]  <(Moderator) Du'An Lightfoot> Nice no Degree. Make a post on
           social and @ me.
```

Though the small size of the audience eliminates the need for @ mentions, which were common in the Twitch chat, we see that users interacting with one another in the audience still use explicit address terms, even when no other chat interaction is occurring. Compare this to the interleaving messages directed at the streamer, which do not appear to require any form of explicit address. This once again supports the presumption of relevance to the video stream. For example, the below example shows how users assume relevance to the broadcast, i.e. the spoken stream and the utterances of the broadcast host, when not explicitly told otherwise.

Example 24

18:09	If you guys don't know Mr. Du'An Lightfoot is a	
18:16	really good friend of mine Even outside of this whole YouTube thing like I talk to this guy - I mean we probably	

18:22	talk once a week right? Like I would say we talk once a week on the phone maybe	
18:28	Maybe once every couple weeks but this dude motivates me, man Du'An Lightfoot motivates me He keeps me going	
18:35	I love him Anyway on our Windows 10 machine we are going to	
18:41	set a static IP address and as you guys can see here this is the network scheme	<No Degree I.T. !> Du'an i finally found a use for my pi4 taught my son how to set it up so he could do school work on it while stuck at home!!
18:46	that we have for our internal network here so we are just gonna go ahead and	
18:52	do 10.0.2. let's do 28 255 uh oh	<Du'An Lightfoot> Love you back man!
18:58	I know 256 doesn't work come on 255.255.0 the default gateway	<No Degree I.T. !> Great Dude for sure
19:06	10.0.0.2 preferred DNS - or is that	
19:11	is that the DNS let's take a look at this before I - I go further I'm gonna spin up this	
19:16	machine real quick I'm all over the place right now	<Michael Anderson> Yeah for sure.

Though the moderator, `Du'An Lightfoot`, has just received compliments from both the broadcaster and another participant, `No Degree I.T. !`, the message `Love you back man!` is unambiguously assumed to be addressed to the broadcaster, while the response to `No Degree I.T. !` is marked with direct address, as other users will assume optimal relevance to the video broadcast, and not another chat message, otherwise.

Conclusion

Both case studies exhibited the phenomena of interest to varying degrees: an expected outcome due to differing technological features of each platform as well as the social norms of each broadcaster's

community. However, both exhibited a similar pattern of disjointed adjacency wherein the difference in the 'tempo' of the spoken and written channels facilitated methods of cross-modal interaction through the non-sequential chaining of relevant conversational exchanges. This appears to be a highly commonplace form of conversational interaction that occurs not only in these case studies, but also broadly in online live streams, as will be discussed in the next section. Both case studies also exhibited cases of independent conversation within the text channel; however, such conversation was likely to be closely related to the topic of the stream or attempt to involve the live streamer's identity, acts, or opinions in some way. Such behavior suggests the importance in looking towards social as well as technological explanations for the ways in which live streams differ from text- or video-only environments of CMC.

V. DISCUSSION

5.1 The nature of cross-modal interaction in live streams

Following previous discussions of the nature of cross-modal discourse in live streams, it is most productive to consider cross-modal interaction in live streams as being composed of two separate phenomena: written to speech interaction and speech to written interaction. Not only do these interactions carry very different norms from one another in the context of live streaming, they also behave differently to norms of 'democratic' video-mediated communication environments (e.g. Zoom conferences), where mode-switching is an option[6]. The ways in which such interactions are carried out in live streams not only tell us about the social structure of live streams and the broadcaster–audience relationship, they also suggest that cross-modal interaction between written and spoken channels may cause spoken channels to behave more like written channels and written channels to behave more like spoken channels with regard to temporal expectations of topic salience and relevance.

Before evaluating the nature of cross-modal interaction between the written and spoken channels, it is important first to characterize how they individually behave in the live stream environment. The case studies consistently showed that the behavior of the text chat in a live stream is not equivalent to the behavior of a standalone text chat as evaluated through my literature review of studies on instant messaging platforms and IRC. Similarly, the behavior of the broadcaster in the video stream is entirely different than that of a single participant in a video conference or other video-mediated communication mode.

Fig. 2 depicts the general multimodal discourse structure of a live stream. The broadcaster in the spoken channel engages in monological moves, paralleling a face-to-face presentation, while users in the written channel also engage in monological moves such as reactionary emojis or spam messages.

[6] Though the broadcast host does have the option to type in chat, this was not observed at all in any of my case studies, and, from my experience viewing live streams, is extremely rare. This is likely a cultural norm, as the audience expects the broadcaster to be engaging in activity, commentary, or interaction through the video broadcast, so stopping to type something in the text chat would detract from such goals.

However, the central interaction occurs between the broadcaster and the audience, constituting either individual interactions or the audience moving as a collective. In the following discussion, I will first characterize the independent speech acts occurring in both the spoken and written channel before describing the nature of interaction across the two.

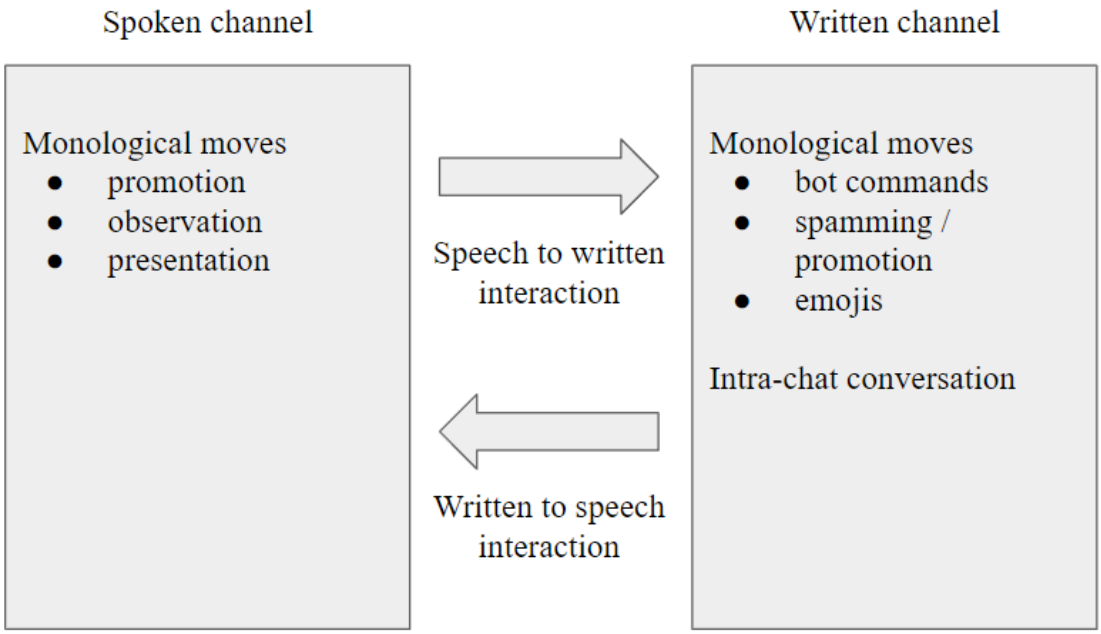


Fig. 2: Representation of discursive moves across the spoken and written channel in live streams. Arrows represent the direction of interaction, pointing from the instigating party to the responding party.

5.1.1 The written channel

As expected, the text chat of live streams exhibits many features similar to those observed in IRC and other quasi-synchronous text-based CMC modes, such as a high degree of disrupted adjacency and low expectations for relevance. However, live streams uniquely feature an even further disjointed text chat due to the one-to-many presentation structure of live streams. Because viewers are responding to the broadcast, and not to one another, adjacent chat messages are almost never assumed to be relevant to each other, unlike in traditional speech or even in quasi-synchronous CMC, in which adjacent relevance is still

expected in slow-moving conversations. Instead, the primary axis of relevance is the 'main floor' of the broadcast, which occurs in a different channel and medium than the text chat. As a result, the text chat displays generally no adjacency standard of traditional discourse, and instead features a community of users independently interacting with the broadcaster.

However, it would not be accurate either to characterize the text chat instead as merely a collection of one-on-one conversations between independent audience members conversing with the broadcaster, who simply happen to be sharing a single channel. The nature of live streams as a group activity means that audience members necessarily interact with one another, whether through independent conversations in the text channel, or through behaving collectively as an audience. Both case studies exhibited intra-audience interaction to varying degrees, which may be attributed to the social settings of each stream, which is further discussed in section 5.1.2. My findings suggest that intra-chat messages were far more common in the Twitch broadcast, which I hypothesize is strongly related not to the technological affordances of the medium, but rather to the familiarity of audience members with one another and the broadcast channel's efforts in curating a distinct social brand and community. When audience members view the live stream as a shared activity, intra-chat messages are more likely to occur than when the broadcast is anticipated to be a chance for audience members to individually interact with the broadcaster.

Nonetheless, given that the majority of utterances in the text chat are in response to the broadcaster or directed at the broadcaster, it logically follows that the text chat is largely composed of monological discourse moves (e.g. spamming reactions, bot commands, or emojis) or incomplete discourse moves (e.g. directing a question or suggestion to the broadcaster that is not responded to or making a statement or response that is not responded to).

However, this is not the only factor contributing to the perceived incoherence of the text chat. Despite the fact that relevance is assumed almost exclusively to the broadcast and the goings-on of the video stream, the interleaving of topics (constructing parallel floors) and fast rate of topic development and decay continues to characterize the text chat of live streams in the same way it does IRC. In IRC and

other quasi-synchronous text-based CMC modes, dialogical moves are frequently interleaved, resulting in multiple topics of conversation being simultaneously sustained through interleaving messages, requiring users to correctly assign each speech act to one of many conversational floors, and frequently leading to ambiguity or needing conversational repair when such assignments are incorrect. The text chat of live streams, as seen in the case studies, continue to exhibit the interleaving of several topics and the generation of parallel floors, even though the majority of text chat utterances do not engage with one another. This is due to the interleaving exchanges occurring with the video stream: when topics are introduced in the video stream, the delay caused by the quasi-synchronous nature of written text means that the written channel does not always follow the same chronological order as the spoken channel, allowing topics to be interleaved.

However, I observe two primary differences between the generation of parallel floors and the interleaving of topics in text-only mediums versus in live streams. First, this interleaving conversation involves the spoken channel and the broadcaster as a conversational interlocutor, with the spoken floor and multiple written floors exerting influence over one another in terms of topic generation and decay in a greater way than interleaving topics in multi-user text chats do. In IRC, for example, though parallel floors may exist in one channel, they are far more distinct from one another than in the chat of a live stream; since there is no common point of focus (the broadcast) exerting influence over all of them, the topics involved can become far less unified than in the text chat of a live stream. Second is that the nature of conversational interleaving in the chat of live streams generally involves interleaving *with the broadcaster* more often than interleaving amongst the chat alone. Though the interleaving of topics in the text chat of broadcasts sometimes involves messages pertaining to multiple topics occurring simultaneously, thus sustaining multiple independent topics simultaneously, the primary way in which the interleaving of topics involves the general range of text chat responses following the spoken stream, which may itself be sustaining multiple topics at one time as it responds to a variety of written utterances. Thus, it is more common to observe multiple messages at once pertaining to a given topic, followed by a chunk of utterances pertaining to a different topic, followed by multiple messages pertaining to the first

topic. Like IRC, this sustains multiple topics simultaneously, but the majority of text messages follow the topic of the video stream, which may alternate between topics, but not at the rate expected on IRC.

5.1.2 The spoken channel

The spoken channel of live streams, when viewed independently, exhibit a wide variety of discursive and monologic acts that make them difficult to characterize. Though previous studies have suggested that larger broadcasts behave more similarly to presentations while smaller ones behave more similarly to conversations, the findings from these case studies have shown that even relatively large broadcasts can have elements of conversation, although the extent to which that conversation occurs and with whom it is conducted varies greatly. Specifically, when it comes to broadcasts that do not feature the host presenting or showcasing a certain ability or activity, such as gaming, drawing, or cooking, but rather broadcasts that center conversation as the activity itself, the spoken channel is thus necessarily interactive with the text channel, and cannot be viewed in isolation (as opposed to, for example, a lecture or keynote presentation, which may have input from the audience, but can be understood standalone).

In the spoken channel, the broadcast host generally carries out one of a set of different speech acts, including monological moves, frequently the sharing or presenting of something; and dialogical moves, such as greeting viewers, asking questions and inviting interaction from viewers, responding to individual utterances from the viewers, and responding to the moves of the viewers as a collective. In broadcasts hosted by a solo streamer, which made up the case studies examined in this study, monological moves constitute the independent speech acts of the spoken channel, similar to how intra-audience conversation and monological speech moves (e.g. bot commands or spamming) constitute the independent speech acts of the written channel, while dialogical moves allow for cross-modal interaction.

5.1.3 Speech to written interaction

Speech to written interaction is the foundation upon which live broadcasts are built. A given utterance by the broadcast host invites responses from the viewers, who may freely respond to the goings-on of the broadcast through the use of the text chat. It is assumed that the host has both a social and financial motivation to do so, as the host's platform benefits from increased viewers and engagement.

Previous studies of video-mediated communication, especially video conferencing, have established that the spoken floor often acts as the 'main stage' of the discourse environment, with spoken utterances holding an expectation of greater importance and salience, while the text chat can behave as a backstage and often includes backchannel. This divide is only further enforced in the live stream environment, where the social structure of the broadcast centers the broadcaster as a performer and their utterances and acts as the central activity of the entire broadcast. This suggests that the discourse environment of the broadcast is made up of technological, cognitive, and social factors.

In video conferencing, it is generally expected that a spoken utterance may incite responses either in the spoken channel or in the text channel, as users are able to freely modulate between the two as they wish. Spoken responses require a much more stringent expectation of conversational maxims such as manner, quantity, and relevance, while written responses need not be as strictly evaluated, as they do not directly interrupt the flow of conversation in the spoken channel. Moreover, multiple possible responses can be crafted to a given utterance, and since users cannot see what other users are typing while crafting their own responses, it is expected that multiple possible responses can occur, which may lead the conversation in many different directions. Only one such direction can be taken at one time in the spoken channel due to the cognitive difficulty of sustaining multiple topics without the permanent record of written communication. However, the text channel can sustain a parallel floor (or multiple), allowing for the branching of topics and thus, the generation of multiple floors of conversation.

In the case of spoken-to-written interaction in live streams, a similar phenomenon is observed, though there are marked differences. Whether such differences are attributed to the technological features

of the broadcast, such as the inability of audience members to speak or engage in the video channel, or to the social setting, such as the microcelebrity status of the broadcaster, remains a question to be further explored.

The first difference is the strict demarcation of the spoken channel as the main stage and the position of the streamer as the conversational moderator. As the streamer has a complete monopoly over the spoken floor, they thus have control over what response(s) are given time on the spoken floor. Moreover, because audience members are limited to the text chat only, all potential conversational topics presented by audience members *must* be made in the text channel. Therefore, written responses to spoken utterances are the primary means by which audience members may attempt to momentarily gain control of the written floor. Thus, audience members must craft their utterances carefully so as to gain the most involvement in the conversation, which is related to having more central status in the broadcast and closeness with the host. This becomes even more apparent when technological factors inflict even stronger restrictions on the ability of audience members to make contributions and gain floor control. For example, a common Twitch setting allows broadcasters to enable a "slow mode" in their broadcast, which limits each user to sending at most one message in a given time interval (e.g. one message per user is allowed every 10 seconds). This is often used in extremely large broadcasts to prevent spamming (the repeated sending of messages, often phatic messages or emojis) and to prevent the chat channel from moving so fast such that users and broadcasters cannot read the messages as they scroll by. This is employed to reduce the 'waterfall of text' effect that occurs when a large amount of users are making monological speech moves, but may also limit interaction between users.

Given that audience members are in many ways competing for control of the written floor and the spoken floor through the broadcaster's moderation, they must make contributions that are relevant and socially appropriate—and fast. Highly delayed messages are almost never responded to by the streamer, while an utterance by the streamer can see responses in the written channel up to several minutes later, despite the spoken floor having long since progressed to another topic. Despite this, explicit topic signaling remains unnecessary in the written channel when relevance is maintained to the topic of the

stream, even with high degrees of delay. An assumption holds that text chat messages, unless labeled otherwise, are responses to a voice channel topic, even if said topic is much older chronologically than a current topic in the text chat. Despite the obvious lag time between the broadcaster uttering a statement, the statement being transmitted to the viewer, and the viewer typing out a response, topics addressed in the spoken stream still dominate written floors over chat-only topics.

Such a phenomenon is in some ways surprising considering the accepted analyses of text- and video-based CMC, many of which are based upon the affordances of the medium. While text leaves a written record, voice and video does not, at least not immediately. Therefore, text-based CMC, when on its own, is generally observed to show a higher degree of disrupted adjacency with responses to topics that are older temporally. On the other hand, video-based communication behaves more like face-to-face conversation, expecting temporal adjacency as it lacks the written record left by text-based CMC. In the cross-modal setting, however, this is no longer the case; the spoken channel becomes less and less 'coherent' with regard to temporal adjacency due to its interaction with the multiple topics and conversational floors going on in the text channel. At the same time, the text channel follows a much more predictable pattern of topics as users assume the spoken channel as the main stage of conversation, although new topics (including those independent from the spoken channel) may still occur.

Given that text-based CMC leaves a written record while video-mediated communication does not, we may be surprised to find that highly delayed responses are more acceptable in speech-to-written interactions than in written-to-speech interactions. While it is true that it requires more time to craft an utterance in the text channel (due to the time required to type out a message, and the fact that the message will not be shown until the entire message has been typed out and sent), that does not account for the fact that text chat messages are more likely to be responses to older topics while spoken messages are not. In fact, because of the written record of the text chat and the much larger population of users making use of it, we might expect that the streamer is more likely and able to read out and thus respond to past utterances in the written channel. However, both case studies found that this was frequently not the case. If a chat message is not responded to immediately, it likely will not be at all. This shows the influence of

the social structure of the live stream over that of the technological structure. The broadcaster, as the sole conversational moderator, must select a topic from a wide range of possible proposals presented by multiple audience members, and therefore places importance upon said topic by bringing it to the main stage of conversation. As the text-based conversation takes a backstage role, its past utterances and topics fall out of immediate salience much more rapidly.

5.1.4 Written to speech interaction

I characterize written to speech interactions as responses the broadcaster makes using speech to an utterance that occurred in the text chat. Unlike speech to written interaction, this form of interaction is utilized only by one participant, the live stream host (assuming a standard setup in which there is only one broadcaster). Because the host must respond to a variety of input messages from the audience, many of which invite conversational interaction through the form of adjacency pair exchanges, the streamer must have a strategy of managing multiple one-on-one and one-to-many conversations that simultaneously occur. Unlike the text chat, in which the indexing of topics is rarely necessary due to the assumption that participants are interacting with the streamer unless stated otherwise, if there are a large number of chat participants, then there is no way for the audience to identify to which message the streamer is responding or which conversational topic they are currently interacting with unless it is communicated explicitly. Thus, live stream hosts typically employ a means of indexing their utterances as responses to an ongoing conversational topic or a specific message made by an audience member.

In my case studies, I observed a few different strategies: 1) using the name of the audience member whom they are interacting with, 2) reading the message in question word for word, and 3) echoing the message while index shifting. There were a few rare examples in which the live streamer directly responded to a chat message without engaging in this kind of conversational signposting; I predict that this may have to do with the size of the audience and thus update speed of the chat impacting how likely the presenter believes the audience will be to understand all given conversational topics as well as

the likelihood that audience members will have read the specific message to which they are responding. This potentially provides an explanation for the more conversation-like behavior of smaller live streams, where the broadcaster needs to engage in fewer separate discourses with individual viewers, and thus requires topic management to a lower degree.

One of the major ways in which written to speech interaction occurred in the live streaming environment was in fact structured quite similarly to the types of interleaving conversation generally observed of IRC and other text-based CMC mediums. Recktenwald (2013) refers to this as 'exchange chaining', and it is a way by which broadcast hosts manage the one-to-many nature of speech-to-written interaction in live streams while still engaging in interaction with individual users, thus still simulating closeness through conversation. Similarly to the way conversations are interleaved in text-based CMC, the text chat involves multiple parallel floors proposed and populated by different users. Thus, the broadcaster may simultaneously engage in multiple topics by interleaving exchanges and adjacency pairs for their end of each respective conversation, resulting in a sequentially incoherent spoken channel.

Though further study is needed to warrant such a claim, it is likely that exchange chaining is not merely a strategy by which the one-to-many nature of live stream conversations is managed, but also a strategy employed by broadcast hosts to productively manage turns given the delay and difference in synchronicity between spoken and written mediums. Both live streams selected for the case studies in this paper were relatively well-attended live streams with many audience members, because I aimed to examine intra-chat conversation and the social relationship between viewers in addition to cross-modal interactions with the broadcast host. Speaking purely anecdotally, however, even in broadcasts in which there may be as few as one active participant in the chat channel, I have continued to observe the presence of exchange chaining as the chat participant responds to new goings-on in the live stream (e.g. a new development in a video game being streamed, or a new topic brought up by the streamer) while the streamer must still respond to past utterances. The exchange chaining model of turn-taking is maintained despite the conversation being one-on-one, with multiple conversational topics interleaved with one another. In this case, it appears that the technological factor of text channel delay as well as the permanent

visibility of past text messages has a greater effect on the timing of turns, perhaps suggesting that the social structure of live streams becomes more and more prominent when the broadcaster's channel gains a greater following and thus develops a community around the broadcasts. However, further study contrasting the behaviors between extremely small streams (i.e. broadcasting to one or two audience members at a time) and relatively larger ones would be needed to further investigate the dynamic between the technological and social factors that impact linguistic features on a live stream.

5.2 Topic flow

Although the spoken channel does not develop and change topics nearly as quickly as the chat, in large part due to being occupied by only one participant, there is still an extremely high degree of topic decay in all aspects of the live stream setting. In both case studies, regardless of how formal the setting or how restricted the intended topics of conversation, high rates of topic development and decay still occurred. Such high rates of topic decay and disrupted adjacency were previously assumed to be uniquely characteristic of text-based CMC. However, the live stream setting, especially a 'Just Chatting' or 'Ask Me Anything' stream, relies on constant interaction between the streamer and users in the chat. These channels therefore exert opposing forces on each other, requiring the streamer to keep up with multiple topics simultaneously, and requiring the audience in the written chat to dynamically shift their responses to remain relevant with the topic(s) currently being selected and addressed by the streamer.

It thus is unproductive to view such environments as consisting of separate but parallel conversational channels that merely occur in a shared space, as is often the way in which video-mediated communication and video conferencing is characterized: users may move from one channel to the other as they wish (though when they choose to do so is still constrained by etiquette and social procedures), and therefore, the two channels hold separate utility in one shared conversational space, with most interaction occurring within-channel, not cross-channel. Cross-channel interaction primarily occurs to facilitate the movement of an ongoing conversation or interaction from one channel to another. This is different in live

streams, where the majority of interaction is necessarily cross-modal due to the inability of users to mode-switch. Therefore, interaction primarily occurs across these channels, not within them.

5.3 Conclusion and proposal for future work

This study aimed to investigate separate case studies of live streams and user behavior in order to not merely classify live streams as behaving more similarly to presentations or conversation, or if so, to written or spoken conversation, but rather descriptively analyze the set of discourse behaviors and strategies taken by users in order to glean more insight about the interplay between the discourse and technological setting. The study specifically looks at broadcasts that feature and encourage conversation, and found that users in the live stream environment, both the broadcast host and audience members, engage in frequent interaction, indicating that live streams can be a productive environment for coherent interaction, motivated by the social goals of each party.

The results also showed that user behavior in live streams is significantly different from that of single-mode forms as well as 'mode-switching' forms such as video conferences and meetings as described in studies of video-mediated communication. While it is still the case that the spoken channel remains the primary 'main stage' of the conversational environment, without the option for multiple users to converse in it, the broadcast host in many ways takes on similar behaviors to the text chat, including fragmented adjacency pairs, exchange chaining with multiple participants, and rapid topic decay, all phenomena that were previously seen as exclusive to text-based CMC. This suggests that the related nature of the text and spoken channels in a live stream setting allow for a dynamic interplay of the two channels, which are more reasonably viewed as fostering interaction between them, rather than two parallel channels that are merely simultaneously ongoing. This push-and-pull interaction causes the spoken channel to take on the timing and turn-taking practices of the text chat, while exerting a unifying influence on the text chat that in many cases limits the extent to which independent topics can generate.

On a broader scale, this study evaluates the interplay between technological features and social factors when it comes to analysis of CMC discourse. While it is possible to attribute much of CMC

discourse to the affordances of the technology, the technological situation is ultimately not the only factor that distinguishes a live stream from a face-to-face conversation or in-person presentation. Instead, the engagement in micro-celebrity and communities of practice on the Internet have given rise to a unique set of social structures that play a major role in the difference in text and video communication behavior in live streams compared to a private video call, meeting, or classroom. The public-facing nature of the live streams enforces a different standard on both politeness and the social goals and gains for each user. The extent of this influence on the linguistic and discourse factors of the live stream setting will require further research to uncover.

Throughout this section, I have posited explanations for the behavior I identified in my case studies, ranging from technological features (such as the broadcaster's monopoly over the spoken channel), cognitive limitations (such as the fact that speech does not have a written record, and thus makes it difficult to engage in multiple topics at once), and social norms (such as the expectation of live stream viewers to behave prosocially, however that is defined for each community). It is difficult to say that one of these explanations contributes more greatly to an analysis of discourse in the live streaming setting, or of CMC as a whole than the others. Instead, I propose that these case studies have begun to show that technological features coupled with cognitive limitations produce unique discourse behaviors, and these discourse behaviors gradually shape social and cultural norms. For example, to hearken back to the notion of Relevance Theory provided in Section 2, "be relevant" is a maxim of cooperation that has an underlying cognitive component: to reduce the cognitive load of dealing with irrelevant contributions. The phenomena observed in this study (such as the expectation of relevance to the video stream, the short nature of independent chat conversations, the broadcaster addressing the audience like a collective, etc.) likely also constitute behavioral norms that developed as a result of technological platforms imposing cognitive limitations, building social norms upon the expectation of reducing cognitive load to maintain coherence. In some ways, this addresses the puzzle pointed out by Herring (2006) regarding why users flock readily to the Internet despite the perceived lack of conversational coherence in CMC interactions. Users engage in behaviors that reduce cognitive load given the technological affordances of each

platform, and expect others to do the same. These behaviors become the cultural norms of online communities, which we now frequently see explicitly codified and enforced in forum and channel guidelines and rules.

Ultimately, though research on the discourse of CMC continues to expand, the realm of multi-modal and cross-modal discourse behaviors continues to be relatively understudied. The results of this study provide a descriptive account of new and relatively undocumented interactional behavior, but also invites further questions on the complex nature of highly multi-modal and multi-participant interaction. As online platforms continue to expand, so too does the complexity of interactions within communities. Though this study focused on solo hosted streams where conversation was strongly invited as a primary activity, further research on this topic is needed to expand upon the proposed framework and analysis to include other forms of interaction that are common in live streams. The diagram in Fig. 3 depicts an extended framework that features other forms of interactions in the live stream setting that have been discussed in past work on the topic (e.g. Recktenwald 2018) and that I have more broadly observed in live streams, such as interactions with other players of the game or interaction with media that is being consumed together on the stream. Such interactions contribute to a multifaceted setting of which this study has only begun to scratch the surface.

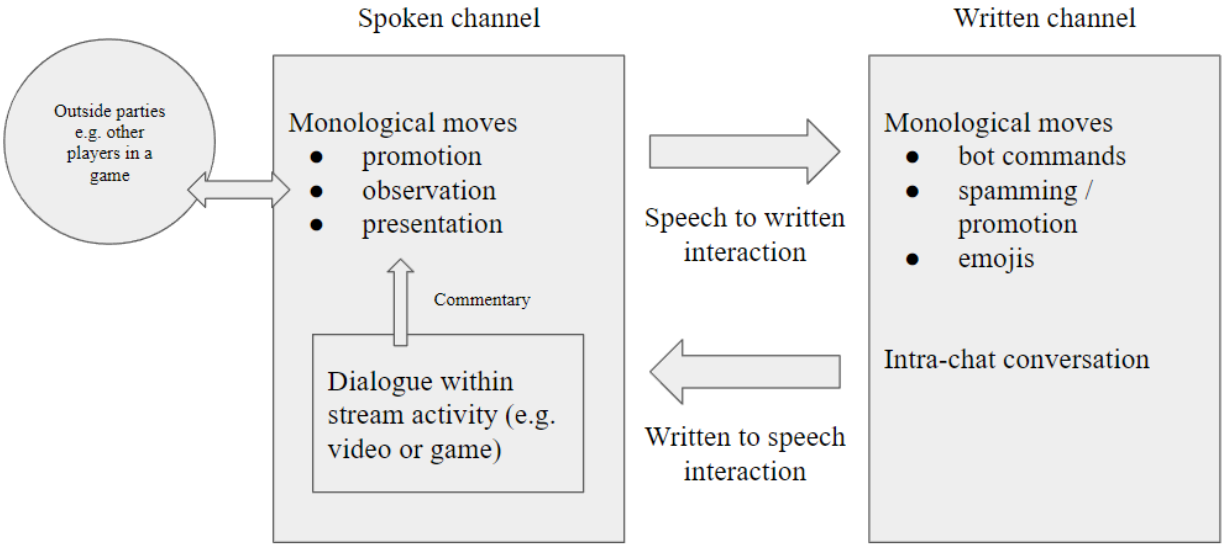


Fig. 3: Expanded representation of discursive moves in live streams

Ultimately, discourse and interaction on the Internet is still a new and developing field of study, and online communities of practice and trends of behavior are continuing and will continue to rapidly change, especially as greater and greater attention is given to the development of digital communication platforms and their integration into entertainment, business, and social interaction. New and further updated research will continue to be necessary to understand these changes as computer-mediated communication, especially multi-modal and cross-modal communication, continues to become even more prevalent as a medium for linguistic interaction.

REFERENCES

Anderson, J. F., Beard, F. K., & Walther, J. B. (2007). Turn-taking and the local management of conversation in a highly simultaneous computer-mediated communication system. *Language@Internet* 7(7).

Bernad-Mechó, E. (2015). A multimodal discourse analysis of linking metadiscursive elements in two Opencourseware lectures (moocs). *Procedia - Social and Behavioral Sciences, 212*, 61-66. doi:10.1016/j.sbspro.2015.11.299

BoringBoredom. (2022). Twitch Chat downloader: Command Line utility that downloads Twitch Chat Logs [Computer software]. Retrieved from https://github.com/BoringBoredom/Twitch-Chat-Downloader

Cherny, L. (1999). Conversation and Community: Chat in a Virtual World. Stanford, CA: CSLI Publications.

Crystal, D. (2011). *Internet linguistics: A student guide*. London: Routledge.

Danet, B. (2001). *Cyberpl@y: Communicating online*. Oxford: Berg.

Durscheid, C., & Frehner, C. (2013). Email communication. In 2440305710 1654440232 S. C. Herring, 2440305711 1654440232 D. Stein, & 2440305712 1654440232 T. Virtanen (Authors), *Pragmatics of computer-mediated communication*. Berlin: De Gruyter Mouton.

Earnshaw, Y. (2017). Navigating turn-taking and conversational repair in an online synchronous course. *Online Learning, 21*(4). doi:10.24059/olj.v21i4.1029

Edelsky, C. (1981). Who's Got the Floor? *Language in Society*, 10(3), 383–421. http://www.jstor.org/stable/4167262

Ferrara, K., Brunner, H., & Whittemore, G. (1991). Interactive written discourse as an emergent register. *Written Communication, 8*(1), 8-34. doi:10.1177/0741088391008001002

Ford, C. E. (2012). Conversation analysis and turn taking. *The Encyclopedia of Applied Linguistics*. doi:10.1002/9781405198431.wbeal0216

Ford, C., Gardner, D., Horgan, L. E., Liu, C., Tsaasan, A. M., Nardi, B., & Rickman, J. (2017). Chat speed op pogchamp: Practices of coherence in massive Twitch chat. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. doi:10.1145/3027063.3052765

Garcia, A. C., & Baker Jacobs, J. (1999). The eyes of the beholder: Understanding the turn-taking system in quasi-synchronous computer-mediated communication. *Research on Language and Social Interaction, 32*(4), 337-367. doi:10.1207/s15327973rls3204_2

Goffman, E. (1959). *The presentation of self in everyday life.* New York: Doubleday.

Herring, S. C. (2006). Interactional coherence in CMC. *Journal of Computer-Mediated Communication, 4*(4). doi:10.1111/j.1083-6101.1999.tb00106.x

Herring, S. C. (2004). Computer-mediated discourse analysis. *Designing for Virtual Communities in the Service of Learning,* 338-376. doi:10.1017/cbo9780511805080.016

Herring, S. C. (2010). Who's Got the Floor in Computer-Mediated Conversation? Edelsky's Gender Patterns Revisited. *Language@Internet, 7*(8).

Herring, S. C. (2013). Discourse in web 2.0: Familiar, reconfigured, and emergent. In 2440361646 1654471605 D. Tannen (Author), *Discourse 2.0: Language and new media*. Washington, DC: Georgetown University Press.

Herring, S. C. (2013). Relevance in computer-mediated conversation. In S. C. Herring, D. Stein, & T. Virtanen (Eds), *Pragmatics of computer-mediated communication*. Berlin: De Gruyter Mouton.

Herring, S. C. (2020). New Frontiers in Interactive Multimodal Communication. In 2440366987 1654474730 A. Georgakopoulou & 2440366988 1654474730 T. Spilioti (Authors), *The Routledge Handbook of Language and Digital Communication*. London ; New York: Routledge, Taylor et Francis Group.

Herring, S. C., Stein, D., & Virtanen, T. (2013). Introduction to the Pragmatics of Computer-Mediated Communication. In *Pragmatics of computer-mediated communication*. Berlin: De Gruyter Mouton.

Jenks, C. J. (2009). When is it appropriate to talk? Managing overlapping talk in multi-participant voice-based chat rooms. *Computer Assisted Language Learning, 22*(1), 19-30. doi:10.1080/09588220802613781

Jenks, C. J., & Brandt, A. (2013). Managing mutual orientation in the absence of physical Copresence: Multiparty voice-based chat room interaction. *Discourse Processes, 50*(4), 227-248. doi:10.1080/0163853x.2013.777561

Lambiase, J. J. (2010). Hanging by a Thread: Topic Development and Death in an Online Discussion of Breaking News. *Language@Internet, 7*(9).

Markman, K. (2005). To send or not to send: Turn construction in computer-mediated chat. In C. Sunakawa, T. Ikeda, S. Finch & M. Shetty (Eds.), *Proceedings of the Twelfth Annual Symposium About Language and Society-Austin* (Vol. 48, pp. 115-124). Austin: Texas Linguistic Forum.

Marwick, A., & boyd, d. (2011). To see and be seen: Celebrity practice on Twitter. *Convergence: The International Journal of Research into New Media Technologies, 17*(2), 139-158. doi:10.1177/1354856510394539

Murugesan, S. (2007). Understanding web 2.0. *IT Professional, 9*(4), 34-41. doi:10.1109/mitp.2007.78

Paolillo, J. C. (2001). Language variation on internet relay chat: A Social Network Approach. *Journal of Sociolinguistics, 5*(2), 180-213. doi:10.1111/1467-9481.00147

Paulus, T., Warren, A., & Lester, J. N. (2016). Applying conversation analysis methods to online talk: A literature review. *Discourse, Context & Media, 12*, 1-10. doi:10.1016/j.dcm.2016.04.001

Recktenwald, D. (2018). *The discourse of online live streaming on Twitch: Communication between conversation and commentary* (Doctoral dissertation). The Hong Kong Polytechnic University.

Rintel, E. S., & Pittam, J. (1997). Strangers in a strange land interaction management on internet relay chat. *Human Communication Research, 23*(4), 507-534. doi:10.1111/j.1468-2958.1997.tb00408.x

Rosenbaun, L. (2016). *Interaction management in recreational video-mediated communication: Participation, multiactivities, and visual playfulness in multiparty Google+ hangouts* (Doctoral dissertation). University of Haifa.

Schegloff, E. A., & Sacks, H. (1973). Opening up closings. *Semiotica, 8*(4). doi:10.1515/semi.1973.8.4.289

Senft, T. M. (2008). *Camgirls: Celebrity & community in the age of Social Networks*. New York: Lang.

Senft, T. M. (2015). Microcelebrity and the branded self. In J. Hartley, J. Burgess, & A. Bruns (Eds), *A companion to New Media Dynamics*. Chichester: Wiley-Blackwell.

Severinson Eklundh, K. (2010). To Quote or Not to Quote: Setting the Context for Computer-Mediated Dialogues. *Language@Internet, 7*(5).

Simpson, J. (2013). Conversational floor in computer-mediated discourse. In S. C. Herring, D. Stein, & T. Virtanen (Eds), *Pragmatics of computer-mediated communication*. Berlin: De Gruyter Mouton.

Sindoni, M. G. (2015). *Spoken and written discourse in online interactions: A multimodal approach*. New York: Routledge.

Sperber, D., & Wilson, D. (2013). *Meaning and relevance*. Cambridge: Cambridge University Press.

Stromer-Galley, J., & Martinson, A. M. (2009). Coherence in political computer-mediated communication: Analyzing topic relevance and drift in Chat. *Discourse & Communication, 3*(2), 195-216. doi:10.1177/1750481309102452

Twitch Games Statistics. TwitchTracker. (n.d.). Retrieved March 20, 2023, from https://twitchtracker.com/statistics/games

Twitch.tv. (n.d.). Company. Retrieved March 20, 2023, from https://www.twitch.tv/p/en/company/

Werry, C. C. (1996). Linguistic and interactional features of Internet Relay Chat. In S. Herring (Ed), *Computer-Mediated Communication: Linguistic, social and cross-cultural perspectives*. Amsterdam: Benjamins.

Xenova. (2023). Chat-downloader: A simple tool used to retrieve chat messages from livestreams, videos, clips and past broadcasts [Computer software]. Retrieved from https://github.com/xenova/chat-downloader

Xu, Y., Vanden Abeele, M., Hou, M., & Antheunis, M. (2022). Do parasocial relationships with micro- and mainstream celebrities differ? an empirical study testing four attributes of the parasocial relationship. *Celebrity Studies,* 1-21. doi:10.1080/19392397.2021.2006730

Yngve, V. (1970). On getting a word in edgewise. In *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society 1970*. University of Chicago.