Computer-Mediated Spoken Interaction: Aspects of Trouble in Multi-Party Chat Rooms

Adam Brandt
Kansai University

Christopher Jenks
City University of Hong Kong

Abstract

Few CMC studies have empirically investigated computer-mediated spoken interaction. This study applies concepts and previous findings from conversation analysis (CA) to the analysis of interactions in multi-party voice-based chat rooms. The research contributes to the current special issue by taking a different look at online trouble talk through the empirical examination of interactional troubles that appear to be unique to this context. The troubles investigated occur during, and/or as a result of, participants’ first verbal contribution to a chat room, next speaker selection, overlapping talk, and speaker identification. The analysis (1) demonstrates how troubles unfold during online spoken communication; (2) shows how interactants respond to, and overcome, troubles; (3) discusses how technology is a mediating factor in trouble talk; and (4) considers how troubles in online spoken communication are similar to, or different from, face-to-face interaction and other voice-only communication settings such as the telephone.

Introduction

In everyday interaction, including computer-mediated communication (CMC), participants routinely display their understanding of each other in, and through, their interactional conduct; they co-construct meaning by, for example, designing their talk to accommodate a recipient, and signaling (verbally or otherwise) comprehension while listening to a speaker. Occasionally, however, participants in talk must deal with various forms of communicative trouble. In face-to-face interaction, such troubles can stem from slips of the tongue, difficulties in hearing, or problems in understanding what has been said. These trouble types extend to CMC contexts (e.g., Herring, 1999; Negretti, 1999), although online environments present additional challenges for interactants (Herring, 2004; Jenks, forthcoming). Forms of trouble that stem from communicating in online settings include background noise, technological failures, and problems resulting from the absence of physical co-presence, such as difficulty in identifying speakers.

In considering and applying conversation analytic concepts to the empirical investigation of “trouble” as a hindrance to interaction in a relatively under-explored CMC setting, we provide a new and unique contribution to the present special issue on “computer-mediated troubled talk.” While past research has explored how interactants deal with troubles in CMC contexts, it has largely been limited to text-based communication, such as email, text-based chat, and discussion forums (e.g., Jacobs & Garcia, 2013; Jepson, 2005; Smith, 2008). These studies show that the affordances and constraints of CMC play an important role in how participants in talk deal with interactional troubles in communication. Although investigations of troubles in text-based
communication have shed light on how computers mediate communication, it is important to extend this understanding to different communication modalities.

The aim of this study is to identify the different types of interactional troubles that occur as a result of communicating in a computer-mediated spoken interaction (CMSI) environment (for an extended discussion of CMSI, see Jenks, forthcoming), and when applicable, to make connections with the existing literature on talk-based troubles in other communicative contexts. The troubles investigated here occur during, and/or as a result of, participants’ first verbal contribution to a chat room, next speaker selection, overlapping talk, and speaker identification. The analysis will (1) demonstrate how troubles unfold during online spoken communication; (2) show how interactants respond to, and overcome, troubles; (3) discuss how technology is a mediating factor in trouble talk; and (4) explore how CMSI troubles are similar to, or different from, face-to-face interaction and other voice-only communicative media, such as the telephone. The methodology used to carry out this investigation is conversation analysis (CA), which can be considered both a method for the analysis of spoken interaction and a theory of interaction itself. The object of analysis for CA researchers is the sequential organization of talk-in-interaction, investigated through detailed examination of recordings of naturally-occurring social interaction, supported by finely-detailed transcripts. As such, this approach is strictly empirically grounded and data-driven.1

Literature Review

Online communication exists in many forms and media (e.g., social media websites, blogs, chat rooms, discussion boards; see Herring, 2004; Thurlow, Lengel & Tomic, 2004), and each communication type possesses many affordances, such as the ability to speak with people who are not physically proximate. Additionally, each communication type brings with it what may be considered communicative constraints, especially when compared to face-to-face spoken interaction (e.g., lack of nonverbal resources, such as embodied conduct).

Decades of research have demonstrated that these affordances and constraints shape online communication in several ways (e.g., Crystal, 2011; Herring, 1999; Hutchby, 2001, 2005; Lotherington & Xu, 2004; Simpson, 2005; Walther, 1997). For example, message persistence in online text-based chat rooms allows interactants to engage in multiple conversational floors, but this facet of communication also constrains discourse cohesion and coherence (Herring, 1999). The mediation of communication by technology is particularly germane to the investigation of troubles in online communication.

Although many CMC researchers are concerned with examining the mediating effects of technology, only a handful of studies have investigated CMSI (e.g., Cziko & Park, 2003; Dourish et al., 1996; Jepson, 2005), and the majority of these have not aimed to uncover the unique features of online spoken communication. For instance, Jepson’s (2005) study is concerned primarily with second language acquisition, and his observations of CMSI are based on statistical analysis of linguistic features. Heins et al. (2007) examine spoken interaction in online tutorials, but the observations made in the study are based on a coding scheme that can be used with any mode of online communication, rather than including features that are unique to CMSI. Numerous other scholars working in language teaching and education have examined

Language@Internet, 10 (2013), article 5. (www.languageatinternet.org, urn:nbn:de:0009-7-34803), ISSN 1860-2029
online spoken communication, and most of these researchers do not analyze CMSI as a highly complex, organized type of interaction, as is evident in the absence of paralinguistic and prosodic features in the transcripts used in these studies (Hampel & Hauck, 2004; Lamy, 2004). Similarly, researchers working in human-computer interaction have investigated online spoken communication, but very few studies have presented a sequential analysis of recorded CMSI data (e.g., Dourish et al., 1996; Huang et al., 2009; Wadley et al., 2007).

To date, it appears that only a few researchers have investigated the interactional character of CMSI. For example, Yanguas (2012) examined second language learners of Spanish engaged in task-based CMSI. Yanguas analysed his experimental group of Skype users in terms of their turn-taking and meaning negotiation, and compared the analytic findings to previous findings for face-to-face and text-based CMC language learning tasks. As with many of the studies mentioned above, Yanguas aims to unpack CMSI with the ultimate goal of applying it effectively to second language teaching and learning. Even fewer studies have explored CMSI outside of second language teaching (although they have examined interactions involving second language speakers)–Jenks, for example, has analysed how speakers in CMSI become acquainted (2009a) and manage overlapping talk (2009b). Additionally, Jenks and Firth (2013) present an overview of how, for example, participants organize turn-taking and identify and recognise one another in such settings. All of these studies, both those with a pedagogical focus and those without, contribute to a more nuanced understanding of how spoken interaction is organised in voice-based CMC.

Because of the nature of this setting as non-proximate spoken interaction mediated by technology, research on spoken interaction involving other technologies is also relevant. In fact, CA as a research approach began in the 1960s with Harvey Sacks’ examination of telephone calls to suicide hotlines (published as Sacks, 1992). Since that time, an enormous body of research has focused on telephone conversations (e.g., Hopper, 1992; Hutchby, 2003; Schegloff, 1968, 1979, 1986) and more recently on mobile telephone interactions (e.g., Arminen, 2005; Arminen & Leinonen, 2006; Hutchby, 2005; Hutchby & Barnett, 2005) and push-to-talk radios (e.g., Szymanski et al., 2006).

However, despite the vast abundance of data examined from such settings, many CA studies do not focus on the data as interaction moderated by technological devices. With regard to telephone-based research, the issue of technology and communication has merely been considered as a convenient “device through which are refracted other phenomena” (Schegloff, 2002, p. 290), leaving the constraining nature of the telephone-call context as an afterthought or a footnote. This has rightly been lamented by some (e.g., Hutchby & Barnett, 2005). In fully understanding interaction, one must consider the setting in which the interaction takes place, not least its affordances, constraints, or what Hutchby and Barnett have also described as its “circumstantial contingencies” (2005, p. 668). This study aims to address a gap in the body of research on CMC, and more specifically CMSI, by considering the troubles encountered by participants in multiparty, voice-based chat rooms in relation to the technological affordances and constraints of such environments.

But what exactly is meant by “trouble?” In this study we align with the conversation analytic operationalization of “trouble.” That is, trouble is anything that the participants themselves orient
to as problematic for, or a hindrance to, the ongoing interaction. In this sense, any problems in speaking, hearing, or understanding are considered forms of trouble in spoken interaction. Orientations to such trouble can be identified when participants put on hold the interactional business at hand in order to first remedy something, for example by asking for an explanation of a question before answering it (orienting to trouble in understanding), by asking for a question to be repeated before answering it (orienting to trouble in hearing), or by reformulating a question themselves before it is answered (orienting to trouble in speaking).

Further detailed discussions of trouble and of the “repair” mechanism used by interactants to identify and remedy trouble can be found in Schegloff (1992, 1997, 2000), Drew (1997), and Sidnell (2006), the last being an excellent introductory overview.

Data

The data analysed for this research were collected from chat rooms that were hosted by Skype. Skype is an online software package that uses voice-over-Internet protocol (VoIP) technology to allow users to call each other for free, as well as to call landline or mobile telephones for a charge. At various times since its creation in 2003 (Skype, 2011a), Skype has offered other services in addition to the “conventional” one-to-one and multi-party video/voice call. One such service, no longer available for unspecified reasons (Skype, 2011b), was online multiparty voice-based chat rooms, called “Skypecasts.” This Skype service is the focus of the present study. Although the Skypecasts service is no longer available, other such chat rooms and multi-party voice-based communication tools exist, for example as offered by Paltalk. We believe that the present findings are relevant to other chat rooms and multi-party voice-based communication of this type, although the extent to which this is the case is an empirical question that deserves further investigation.

Skypecasts (referred to in this article henceforth as “chat rooms”) could be set up by any Skype user who, upon creating a chat room, was known as that room’s “host.” Once created, the chat rooms were listed on the Skype website and searchable according to title, description, keywords, and start time. Many of the chat rooms were themed around particular discussion topics, such as sport, religion, and politics. Users who were logged into Skype were able to search through the listings to see upcoming Skypecasts and also join any ongoing chat room.

Skypecast chat rooms afforded three levels of participation: the “listening room,” the “waiting room,” and the “speaking room.” Upon clicking the “Join this Skypecast” link, participants joined the “listening room.” As the name suggests, chat room members were only able to listen to the chat room discussion at this level of participation. Those wishing to contribute verbally to the discussion were required to click on the “Ask to talk” link (as shown on the right hand side of Figure 1). Upon clicking that link, participants were taken to the “waiting room,” from where only the host could promote them to the “speaking room.” However, participants in any of the three levels of participation were able to send private written instant messages to any other individual in the chat room. All Skypecast participants were also able to see a live listing of active chat room members and their current level of participation.
Figure 1 shows a screenshot of the Skype user interface during participation in a Skypecast chat room. As can be seen in the figure, the user who set up the chat room is highlighted in bold and indicated as the host. Although the host plays an important role in how participation in a chat room is organized, in that s/he determines who is permitted into the speaking room, on an interactional level the host has no greater or lesser role than any of the other participants. For example, in dealing with any forms of interactional trouble, hosts are never oriented to, either by themselves or by interlocutors, as authorities in trouble resolution.

![Figure 1. Screenshot of Skype user interface during a live Skypecast](image)

Skypecast chat rooms had no facility for video, and so participants were not able to see one another. Similarly, although chat room members were listed on the Skype user interface, the interface was not able to indicate which specific participant was speaking at any particular time.³

This study includes two corpora of data. The first was recorded with the help of a research assistant between 30 April and 21 May 2007. The second set was recorded by the first author between 28 May and 22 June 2008. A total of 32 recordings from 32 different chat rooms were made, resulting in almost 24 hours of recorded interaction. Individual Skypecast recordings ranged from 15 minutes to just over two hours in length. Skypecast listings were searched, relevant chat rooms were located, and the start dates and times noted. Chat rooms were selected...
for recording from the Skypecast listings if the title or description made reference to the practice or improvement of English as a second language (this was a research interest present prior to data collection, but it was later set aside to examine the interactional character of the setting more broadly).

Upon joining the listening room of a Skypecast, the other participants were informed by both written and spoken message of the researchers’ desire to record the room for research purposes. When consent was given, recording began.

It is difficult to know the demographic details of the participants in the recordings. However, based on voices and personal information offered during the talk, participant ages appeared to range from the late teens (17) to retirement ages (early 60s and older). Participants’ ethnic and national origins varied widely, with chat room members stating that they were from countries in Europe, the Middle East, South America, North America, Africa, and Asia. Although proficiency in English varied quite broadly, the vast majority of interaction was conducted in English. Unlike in some Skypcasts, there was no set discussion topic in the rooms recorded; the only ostensible purpose for entering these rooms was to participate in conversations conducted in English.

Recorded data were transcribed using the conventions of CA (see Appendix), as pioneered by Gail Jefferson (e.g., 1983a, 1985, 1996, 2004). These transcriptions are deemed necessary support to the primary data, the audio recordings (e.g., ten Have, 2007; Hutchby & Wooffitt, 2005; Schegloff, 2007; Sidnell, 2010). In line with the principles and traditions of CA, the analysts followed a process of “unmotivated looking,” whereby we looked through the data with no a priori intentions (other than looking at interactional features of the setting) and observed what appeared interesting. Many of the phenomena discussed in the present study were initially observed and noted at that stage. After the initial general observations, an agreement was reached on which elements of interactional trouble to focus upon, and collections of examples of each interactional phenomenon were assembled. Following conversation analytic traditions, each individual case was analyzed in its own right before general trends were noted. In the analysis presented below, prototypical examples from each of the analytic foci are presented.

**Analysis**

In this section, we present analyses of three distinct, albeit related, forms of trouble that participants in multi-party voice-based chat rooms are faced with. These forms of potential trouble are: (1) being unable to join in the ongoing talk; (2) knowing if, and if so when, you are expected to be the next speaker; and (3) identifying who your interlocutor is. These three aspects form sub-sections in this analysis section. In each of the three instances, we provide a number of exemplifying cases for the form of trouble in question and provide a detailed line-by-line sequential analysis for each case, utilizing the principles and prior findings of CA. It should be noted that these forms of trouble are not exclusive; there are many other forms of trouble in such chat rooms (such as technical difficulties), but they have been chosen because we believe them to be the most important given the tools afforded by a CA approach, through which the fine details of social interaction can be examined and analysed from a participant-relevant perspective.
**Trouble Type 1: New Participants Joining in the Talk**

In many forms of spoken communication that lack physical co-presence (e.g., telephone and mobile phone communication), there is typically a fixed number of interactants (usually two) who are participating from the onset of the interaction until its completion. However, one of the features of Skypcasts is that participants are able to join a chat room at any time, provided the host clicks the permission-granting link.

Despite this situation, participants who have been given permission to talk by the host are not automatically entitled/able to join in the ongoing talk. In fact, on many occasions, a new participant’s first utterance is not responded to by the other chat room members, as in excerpt 1 below. This may be understood as one form of potential “trouble” which participants new to a chat room face; how to construct and time their first turns at talk, such that they are able to join in the ongoing talk, and initiate a new participation framework (Goffman, 1963; Goodwin, 2000) of which they are a part. This phenomenon is investigated in more detail, with a specific focus on when it is “successful” or unproblematic, in Jenks and Brandt (forthcoming).

As excerpt 1 begins, Cecilgee, Jan, and others are getting acquainted and negotiating an appropriate topic of discussion for the chat room. The names listed are pseudonyms of the participants’ user IDs provided by the authors. Since participants logged into the Skypcasts with their Skype accounts, their Skype usernames were available for others to see. Occasionally, participants self-identified with their “real” names; we also replaced those names with pseudonyms.6

**EXCERPT 1**

```
01 Cecilgee: let's talk about our hobbies
02 Jess: about [what ]
03 Sammy: [%okay]
04 Newbie: (hello [everyone] )
05 Jan: [about ↑what]
06 Cecilgee: ↓hobbies you know
07 Jeff: =↑(karate)
08 (0.3)
09 (0.6)
10 (0.7)
11 Jan: ↑hobbies=
12 Cecilgee: yeah right.
```

In the above excerpt, Newbie makes his first verbal contribution to the chat room at line 06. Although this was not completely intelligible on the recording, it appears to takes the form of an open greeting to the other members of the chat room (“hello everyone”). Despite this turn being designed such that any other participant could provide a return greeting, no one does. This may be due, at least in part, to the sequential location of Newbie’s turn, as discussed below.

At line 01, Cecilgee proposes a new topic of discussion for the chat room, namely “hobbies.” After a relatively lengthy silence (1.5 seconds, line 02), Jessica indicates that she has either not heard or understood the proposal by initiating repair (line 03). In overlap with this, Sammy
indicates that he is willing to go along with the topic suggestion, with an agreement token “okay” (line 04). The overlap of Jessica and Sammy’s turns is followed by a silence of 0.8 second. Jenks (2009a) has demonstrated how overlapping talk in this interactional setting can often be followed by silence. However, Jessica’s need for a repetition or clarification of the source of her trouble remains unresolved. It is at this point that Newbie comes in with his open greeting. This greeting is produced partly in overlap with Jan’s turn at line 07, in which she appears to pursue repair of Cecilgee’s still-to-be-repaired turn at line 01. What follows in lines 09-14 is a repair and confirmation sequence between Cecilgee and Jan. As a result of the sequential placement of this repair sequence, Newbie’s greeting is not responded to, and so the participation framework is not altered in order to accommodate him.

Excerpt 2 below is a similar case, although it would be reasonable to believe that the new participant in this case, Newbie2, has timed his first contribution more appropriately. In this excerpt, Jan, Melody, and another participant are getting acquainted. In the moments preceding line 01, they have been discussing South Korea, which Jan has just announced is her home country.

EXCERPT 2

01 Melody: ye:s (.) an::d, (0.3) it's
02 really near china right†
03 (1.6)
04 Jan: yeah it's near china too
05 (2.0)
06 Newbie2: ↑hello:=
07 Melody: =yes I'm chinese hhh.
08 (0.4)
09 Jan: ahh are you chinese† okay nice
10 to meet you
11 (0.8)
12 Melody: yes (0.2) nice to meet you too:
13 (1.3)
14 Jan: so are you in china† right now†

At line 01, Melody displays some knowledge regarding South Korea—that it is geographically proximate to China. She ends her turn with “right?” (line 2), which both projects (or makes relevant) a response from another speaker and also displays the strength of her belief (in that the construction of the turn also strongly projects a confirmation). At line 04, Jan confirms Melody’s statement. This confirmation is followed by 2.0 seconds of silence (line 05) before Newbie2 makes his first contribution, again with an open greeting, “↑hello:” (line 06).

However, the end of this greeting is latched with another turn from Melody (line 07), with which she contributes some of her own biographical information to the ongoing talk, i.e., that she is from China. After another short silence (line 08), Jan responds to this, and a “nice-to-meet-you” exchange takes place (lines 09-12). Again, the new participant’s first contribution is not taken up by any of the other participants.

The third example of this phenomenon suggests that avoiding overlap might be more important than sequential timing if a participant wishes to be responded to. Excerpt 3 begins as Jan describes to Nelson which part of the UK she is currently residing in:
EXCERPT 3

01 Jan: u:::h (0.5) newcastle it's north
02 (0.4) north:, (0.3) north area
03 (2.1)
04 Nelson: oka::y
05 (4.2)
→ 06 Newbie3: hell[o ]
07 Nelson: [why] so quiet (*)
08 (2.8)
09 Cecilgee: do you know movies=

At lines 01-02, Jan names the city she is living in and explains in which part of the country it is located. A short silence follows (line 03) before Nelson demonstrates uptake (line 04). The 4.2 seconds of silence that follow this (line 05) would suggest that the sequence is closed. As such, it would appear an ideal time for a new participant to make his or her first verbal contribution.

However, when Newbie3 comes in with “hello” (line 06), the end of his turn is produced in overlap with Nelson’s meta-comment about the lack of talk in the room. In the following 2.8 seconds (line 08), no participant returns the greeting, nor does Newbie3 pursue a response (for example, by repeating his greeting). Instead, another participant, Cecilgee, initiates a new topic (line 09). The talk then moves accordingly to that subject, and Newbie3 does not speak again.

This might suggest that existing participants leave the onus on a newly-joined participant to find a means of joining in. In support of this, there are even cases where the new participant’s first verbal contribution is responded to, but he or she still does not become involved in the talk proper. Excerpt 4 below shows one such instance; this exchange took place only minutes after excerpt 3, but it appears to involve a different new participant (at least the voices sounded markedly dissimilar).

EXCERPT 4

01 Jan: i don’t know↑ (. ) i eh- i never
02 heard of that movie (0.3) (sorry)
03 (2.5)
04
05 Newbie4: hi↑
06 (1.2)
07 Jan: hi:::↑
08 (6.5)
09 Jan: so andr[e what are you-]
08 ?: [ what what ]
09 (. )
10 Jan: u:::h what do you do
11 (0.9)
12 Jan: what’s your job↓
13 (0.2)
14 Cecilgee:.hhh hhh (. ) what I am doing
15 now?
16 (0.6)
17 Jan: yes:↓
18 (0.4)
Jan responds to a question about a specific movie by noting that she has never heard of it (lines 01-02). After a 2.5 second silence (line 03), Newbie4 takes the opportunity to offer an open greeting token (line 03), the uprising intonation of which may be used as a further means of seeking response (Couper-Kuhlen & Selting, 1996). Unlike the previous excerpts in this section, this greeting is responded to, by Jan (line 05).

Previous research on greeting exchanges suggests that the initial greeter has the right to speak again once his or her greeting has been returned (Schegloff, 1986). However, in this case, Newbie4 does not speak again. One can speculate that, as a new member of the chat room, Newbie4 was waiting for a welcome or an invitation to introduce himself, etc. However, in the ensuing 6.5 seconds (line 06), no participant speaks. After this silence, Jan self-selects and initiates a new topic of discussion, marking it as such (with a turn initial “so”; Bolden, 2009) and directing it to a particular interlocutor (“Cecilgee”). As with the previous excerpts, Newbie4 does not participate verbally again.

As the analyses in this section have shown, participants admitted into a multi-party chat room do not automatically gain the right to speak, or rather, to be responded to. Depending upon the sequential position of their first verbal contribution, as well as their timing vis-à-vis other participants’ talk, newly-entered chat room members may have trouble becoming an active part of the ongoing talk. As will be seen in the following analysis section, simultaneous, or overlapping, talk is not only a potential trouble-point for newly-joining chat room members, but it is also oriented to by all as something to be avoided.

**Trouble Type 2: Speaking One-at-a-Time and Knowing Who is Expected to Speak Next**

Once a participant has entered the speaking room and has become a ratified part of the spoken interaction, other forms of potential trouble still exist. One form of potential trouble for all participants is speaking simultaneously. One of the most basic principles in the organization of social interaction is that there is a strong preference for “one-speaker-at-a-time” (Sacks, Schegloff, & Jefferson, 1974). In face-to-face multiparty talk, participants are normally able to draw upon physical resources, such as gaze and embodied conduct, to determine who is speaking, who is about to speak, and also to project when the current speaker is about to stop. However, such resources are not available in the Skypecast chat rooms.

Jenks (2009a) has shown in some detail how participants in Skypecast chat rooms use silence as a resource to manage overlapping, simultaneous talk in multiparty contexts. Excerpt 5 below also shows how overlapping talk can be managed in this setting, albeit on an occasion in which there are only two participants in the chat room. The example is included for illustrative purposes and because it would appear to be connected to the next form of trouble discussed further below.

**EXCERPT 5**

01 Jan: nice to $meet you=hh$
02 (0.9)
03 Sayaka: nice to meet you too:
04 (0.2)
05 Jan: .hhh i'm [ kor- ]
06 Sayaka: [(****)]
07 (0.4)
Having just introduced themselves to one another, and after exchanging pleasantries at lines 01-03, Jan and Sayaka are about to launch into their first topic. At line 05, Jan is apparently about to announce her nationality, which is a typical resource used in the initiation of topic. However, the end of her turn is produced in overlap with something inaudible uttered by Sayaka (line 06).

This overlap results in a short silence (line 10) before Jan invites Sayaka to repeat her turn through the use of an open case repair initiator (Drew, 1997) at line 08. This turn, too, is produced in overlap with something said by Sayaka (line 09). Again, a short silence follows (line 10), before Sayaka quietly and quickly appears to make a meta-comment on who should speak next (“you” at line 11). Despite this, neither participant speaks in the next 3.3 seconds. There is an apparent reluctance on the part of both participants to be the next speaker, which extends for another 7 lines until Sayaka finally initiates a new topic at line 20.

As this example shows, participants are cognizant of the risk of speaking simultaneously and the potential trouble that this can cause. For this reason, they can at times appear to be reluctant to speak, unless they are certain that they have the right (or obligation) to do so. This has implications for participants in responding to a previous turn that may not be clearly directed at them. Excerpt 6 shows a simple case in point.

**EXCERPT 6**

<table>
<thead>
<tr>
<th>Line</th>
<th>Transcript</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Sid: so, can you introduce yourself</td>
</tr>
<tr>
<td>02</td>
<td>(1.6)</td>
</tr>
<tr>
<td>03</td>
<td>Cam: eh?</td>
</tr>
<tr>
<td>04</td>
<td>(0.4)</td>
</tr>
<tr>
<td>05</td>
<td>Cam: me↑</td>
</tr>
<tr>
<td>06</td>
<td>(0.9)</td>
</tr>
<tr>
<td>07</td>
<td>Sid: yejr ah↑</td>
</tr>
<tr>
<td>08</td>
<td>(0.8)</td>
</tr>
</tbody>
</table>

At line 01, Sid invites the new chat room member to introduce himself to the other participants. Despite being a new addition to the room, and so perhaps the most likely to be expected to introduce himself, Cam initiates repair at line 03 (“eh?”). When this open-case repair initiation does not produce a response, Cam initiates repair again, albeit of a more specific form; the check at line 05 (“me↑”) appears to seek confirmation that he is the intended recipient of the introduction request. Sid duly confirms this, and the introduction subsequently takes place. This example shows the occasional doubt that participants in these chat rooms can display in speaking
next. This may be in part due to the lack of non-verbal resources normally employed in next-speaker selection; further, it may be indicative of participants’ cognizance of the likelihood of trouble which can ensue following simultaneous, overlapping talk.

In the following excerpt, which shows another example of such a series of events, there appears to be no specific intended recipient, which also leads to ambiguity about who should respond. In this excerpt, Castro, Chanel, Heraldo and others are getting acquainted for the first time.

**EXCERPT 7**

```
01  Castro:  i ↓see=
02  Chanel:  =yes=
03  Heraldo:  =yeah ↓sometimes
04          (6.4)
05  Castro:  so:::: what do you guys do for
06          ↓living
07          (0.4)
08  Castro:  >what do you ↑do<
09          (1.7)
10  Castro:  are you:::: (0.7) are you
11          working,
12          (1.5)
13  Castro:  or are you a ↓student
14          (3.2)
15  Heraldo:  ↑me
16          (1.0)
17  Castro:  mm-hmm.
18          (0.6)
19  Castro:  yeah i- [i mean]=
20  Heraldo:  [u::h=a]
21  Castro:  =both (0.2) both of you
22          (1.3)
23  Heraldo:  u::h i::::'m a:: mechanical
24          ↓engineer
```

The previous topic of discussion is brought to a close by Castro, Chanel, and Heraldo at lines 1-3; this is followed by a lengthy silence (6.4 seconds, line 04). Castro then launches a new topic at line 05, marking it as such with an elongated “so::::” Castro’s subsequent question “what do you guys do for a living” appears to have multiple recipients, but as has been seen, multiple simultaneous speakers are treated as problematic.

Accordingly, there is no response to his question in the following 0.4 seconds, and Castro reformulates his question (line 08), which affords his intended respondents a second opportunity to provide an answer. There is still no response from either Chanel or Heraldo in the following 1.7 seconds (line 09), and so Castro reformulates his question again, this time with a candidate answer built into it: “are you:::: (0.7) are you working” (lines 10-11). At this point, a confirmation or rejection would be sufficient to respond to Castro’s turn. However, most probably because there is still not a clear intended recipient, no such response occurs in the following 1.5 seconds. At line 13, then, Castro produces a possible alternative (“or are you a ↓student”). Again, in the next 3.2 seconds, no response arrives. Finally, at line 15, Heraldo responds to Castro and, as in the previous example, seeks confirmation that he is the intended
recipient (‘‘†me’’). Note that Castro offers a non-committal confirmation at line 17 (‘‘mm-hmm’’) and then explicates that his question was intended for both Heraldo and Chanel at line 19-21 (‘‘yeah i- [i mean] both (0.2) both of you’’).

Again, this interactional ‘‘trouble’’ appears to be a result of the ambiguity in next-speaker selection, coupled with the lack of physical resources available for the participants to draw upon. Were this a face-to-face interaction, Castro could have indicated a particular intended recipient or, alternatively, Heraldo and Chanel could have negotiated non-verbally who was to respond first.

In this section, analysis of various excerpts has shown that simultaneous, or overlapping, talk is treated by participants as problematic, as can be seen by their attempts to avoid it. Additionally, there are occasions when participants appear to have trouble in identifying who is the intended recipient of a turn at talk, such as a question. As has been suggested, these forms of trouble appear to be a consequence of interacting in a virtual ‘‘room’’ full of voices, without the ability to draw upon other interactional resources such as one’s own, and/or one’s interlocutor’s, body. In the final analysis sub-section, a third phenomenon related to this is discussed.

**Trouble Type 3: Identifying Interlocutors**

The final form of participants’ trouble that we analyzed for the present study is that of identifying who is speaking. Participants in the chat rooms regularly orient to speaker identification as important and often will put on hold the ongoing talk in order to establish the identity of their interlocutor, as will be shown.

As discussed in the first analytic section, new chat room participants may have trouble joining in the ongoing talk. When participants are successfully responded to, there is a strong preference for self-identification before talk can proceed with their participation. This is shown in excerpt 8 below.

**EXCERPT 8**

```
01 Diablo: hello guys
02 (0.6)
03 Amaris: [hello]
04 Aramis: [hello]
05 (1.4)
06 Sara: hello[::
07 Amaris: [who is] talking
08 Sara: diablo
09 Diablo: yeah this is diablo speaking
10 how are [you doing]
11 Amaris: [where are] you from[:
12 (0.5)
13 Sara: welcome [diablo]=
14 Diablo: [ fine ]=
15 Amaris: =just good just good (.) what about
16 you
```

At line 01, Diablo makes his first verbal contribution to the chat room. After a short pause, three of the existing chat room members respond (lines 03, 04, and 06). Amaris seeks identification of
a speaker by asking “who is talking” at line 07. Note that even though this question is produced in overlap with Sara’s turn at line 06, and even though both Amaris and Aramis have spoken more recently than Diablo, Amaris’s question is seen to be pertaining to the speaker at line 01. Sara immediately responds to Amaris’s question with “Diablo” (line 08), which the speaker himself aligns with immediately after: “yeah this is diablo speaking” (line 09).

These participants orient to the norm of identification upon joining the chat room. Note also that once he has self-introduced, Diablo is welcomed by the host, Sara, at line 13. It is also noteworthy that prior to this (at line 11) Amaris asks Diablo to state where he is from. This is not responded to, and the lack of response is not treated as problematic by any of the participants. In other words, this excerpt would suggest that identification is a legitimate reason to put other business on hold, but other biographical queries may not be. In this excerpt, we can see a greetings sequence (lines 01-06) followed by a request for identification (line 07) and a subsequent identification (lines 08-09). As such, the sequence is interactionally similar to those found in, for example, telephone interactions – although in this case, the to-be-identified interactant is joining in an ongoing conversation, and the identification sequence involves more than two parties.

The next excerpt shows a similar series of events. However, in this instance, none of the participants have recently joined the chat room, and all have been talking to one another for a short time. Accordingly, the difference between this context and that of one-to-one voice-only interactional settings, such as the telephone, is more clearly marked: Even though the interaction is ensuing, the participants put the conversation on hold in order to re-establish who is present and talking.

**EXCERPT 9**

```
01  Cecilgee: (3.2) .hhh let's change subject
02  Cecilgee: (3.9) let's (0.3) talk about the music we are listening to
03  Cecilgee: (1.6) sen:::dero↑sa (0.2) right↑
04  Jan: (0.2) okay hhh. (0.3) >okay
05  Cecilgee: (1.0) Cecilgee yes
06  Cecilgee: (0.5) so, (0.2) what kind of music do you like cecil
```

After a short silence at line 01, and with the previous talk having come to an apparent end, Cecilgee suggests starting a new topic of discussion (line 02). After almost 4 seconds, during which time none of Cecilgee’s interlocutors respond, he proposes a specific topic for discussion, “let’s (0.3) talk about the music we are listening to” (lines 04-05). Instead of responding to Cecilgee’s suggestion, however, Jan self-selects and checks the identity of the just prior speaker, producing a candidate name, “so you are sen:::dero↑sa (0.2) right↑” (lines 07-09). (Cecilgee’s
username is listed as “Senderosa,” while his “real” name is stated on his profile as Cecilgee). The elongated pronunciation of the name would suggest that Jan is reading this from the Skype user interface, which, as mentioned earlier, lists all chat room participants. Cecilgee confirms that he is the participant who Jan is referring to at line 10, and then Jan receipts this at lines 12-13. After this sequence, at line 15, Jan aligns with Cecilgee’s topic suggestion by posing a music-related question to him and launching the new discussion.

What is particularly noteworthy about this sequence is that the initial name proposed by Jan (“Senderosa”) and the name confirmed by Cecilgee and receipted by Jan (“Cecilgee”) are different, and neither of the participants orient to this as problematic. This suggests that it is not the accurate naming of a participant that is essential, but rather simply identifying which voice matches with which username on the list of chat room participants.

The final excerpt in this section is similar to excerpt 9, in that the participants are engaged in an ongoing discussion which is put on hold in order to identify a speaker. This example is an extension of excerpt 8, where Jan, Jems, and Cory are getting acquainted.

**EXCERPT 10**

01 Jan: hey↑ lee↑=
02 Jems: =[yeah (**)]=
03 Cory: [ hahu↑ ]=
04 Jan: =do↑ in korea↑
05 (0.3)
06 Jan: [ (**) ]
07 Jems: [ yeah↑↑ ]
08 (0.4)
09 Jan: ar- are you a student↑ or:: are you working↓
10 (1.6)
11 Jems: me↑
12 (0.4)
13 Jan: yes↓
14 (4.6)
15 Jems: >who is spea[king (*)] =
16 ?: >=(let’s do it)/(that’s true)<
17 Jems: =jan↑
18 (0.6)
19 Jan: yeah↑ (. ) I just [ asked you. ]
20 Jems: [ yeah yeah yeah ]
21 (0.3)
22 Jan: whether [ you]=
23 Jems: [ yah. ]
24 Jan: =work or y’know↓ (0.2) whether you study↓=
25 Jems: =a::h↑ yeah yeah

Lines 01-08 are included to show the build-up to the sequence of interest, which begins with Jan’s question at line 09 (“ar- are you a student↑ or:: are you working↓”). Jems checks at line 11 whether he is the intended recipient of the question, and Jan confirms this at line 13. However, before answering the question, Jems checks who it is who has asked the question; “>who is speak[king” (line 15). Note that, although there is usually a strong preference to have a question answered, Jems puts that business on hold in order to identify his interlocutor.
Jems then provides a candidate answer for his own question, i.e., that it is Jan who he is speaking with (line 17), which Jan confirms is the case (line 19). Jan then continues to talk, re-establishing through lines 19, 22, and 24 exactly what her question is. Throughout this, Jems displays that he does not require the question to be repeated, overlapping with multiple “yeah” tokens (lines 20, 23, 25). This would seem to be further evidence that Jems has understood the question, but had prioritized identifying his interlocutor before providing an answer to it.

As has been shown in this section, participants treat non-identification of an interlocutor as troublesome, by halting the ongoing talk in order to identify who it is that they are talking with. Trouble in connecting a voice to a name (be that a username or a “real” name) is something that participants in such chat rooms are always at risk of encountering.

Discussion

Through our analyses, we have demonstrated some of the troubles that participants encounter during their interactions in voice-based multi-party chat rooms. The troubles we have examined are those that occur when (1) a new interactant tries to join in the ongoing talk; (2) participants speak simultaneously or do not know who ought to be the next speaker; and (3) participants do not know who they are speaking to.

When new interactants enter a voice-based chat room, it is apparent that the timing of their first verbal contribution is vital if they are to join in the ongoing talk successfully. Our analysis showed how first verbal contributions may not be responded to if they are produced in overlap or if they are produced at a time when another speaker has been projected to talk (such as by being asked a question). In such cases, the newly-joining interactant’s presence may go unacknowledged. In fact, other analysis suggests that newly-joining participants’ best means of joining in the chat room talk is to wait for the ongoing talk to come to a halt and/or wait to be introduced by an existing chat room member (Jenks & Brandt, forthcoming).

This trouble in becoming part of the “participation framework” (Goodwin, 2000) is in no small part due to the participants’ lack of physical co-presence. When groups of speakers who share some physical environment are approached by another participant, eye gaze and bodily orientation(s) can adjust to afford the participant entry into the participation (Goffman, 1963) before a verbal contribution is even made. In this environment, even when participants have the visual resources to note a new participant (i.e., the user interface with its list of existing chat room members), there is no guarantee that these resources will be utilized (or even noticed). Further, even when chat rooms members are listed as present in the speaking room, they are not necessarily expected to speak (Jenks & Brandt, forthcoming). In this sense, the lack of a physical presence makes it somewhat more troublesome for a new chat room member to engage in talk.

Additionally, unlike in other forms of multiparty online chat such as synchronous text-based chat rooms, a participant’s contribution in the voice-based chat rooms is ephemeral. Once a verbal contribution has been made, it leaves behind no trace (unlike a written contribution, which is at least semi-permanent and so can be revisited more than a verbal contribution can). In this sense, too, voice-based chat rooms can be considered potentially troublesome.
The absence of physical co-presence, coupled with the fact that verbal contributions occupy a fixed temporal location, are also factors in the second type of trouble we addressed: simultaneous talk and problems in next-speaker selection. Lerner (2003) and Goodwin (1980, 1981) have demonstrated the important role that eye gaze plays in next-speaker selection. In the absence of this, participants have fewer resources with which to determine who should speak next when no next speaker has been explicitly selected in the previous turn at talk (such as by a question being directed at a specific party) and who should resume/continue speaking when two or more participants begin to speak simultaneously. These forms of interactional trouble are treated as a hindrance to comprehensible and meaningful communication.

The participants also orient to the problematic nature of not having a name associated with a voice. As was explicated in the final analytic section, participants will put the ongoing talk on hold in order to ascertain to whom they are speaking. This would appear to be unique to voice-based multi-party chat rooms. In physically co-present interaction, connecting a voice to a specific interlocutor is not an issue. Similarly, in other voice-only settings such as telephone calls the identities of interlocutors are almost always ascertained at the beginning of an interaction (e.g., Schegloff, 1968, 1979), and identification is built into the openings of mobile telephone-based spoken interactions (Arminen, 2005; Hutchby & Barnett, 2005). Again, the unique nature of this setting appears to bring with it unique interactional obstacles, which the participants have to deal with in situ and ongoingly.

By focusing on troubles in the present study, we are not implying that Skypecasts in particular, or CMSI in general, is an inherently and exclusively troublesome mode of interaction; interactional troubles occur, and are of analytic interest, in all interactional settings. The setting of voice-based multi-party chat rooms brings with it many affordances and opportunities for new forms of conversations. However, this setting is relatively new, and many users are still learning how to interact in such environments in order to maximize successful communication. Additionally, in examining the interactional troubles that participants may face, we note that the participants are able to deal successfully with these obstacles and have meaningful, comprehensible interactions.

Finally, in focusing on these issues, we have foregrounded interactional trouble over, say, technological troubles, which are also prone to occur in interactional settings mediated by computers. The trouble types that we have focused on are not necessarily as obvious as software crashes or static noise interferences; in fact, troubles such as those addressed here are often only identifiable and understandable to researchers through the lens of the micro-analysis of social interaction. As such, our study adds further weight to the evidence in favor of utilizing such methodological resources in the analysis of CMC. The analysis also adds to understanding the interactional troubles that participants in these settings may encounter.

Notes

1. Readers unfamiliar with the history or principles of CA are advised to read introductory texts by ten Have (2007), Hutchby and Woofitt (2005), and Schegloff (2007). Additionally, the early research papers by Sacks, Schegloff, and Jefferson (1974) and Schegloff, Jefferson, and Sacks (1977) provide a flavor of the analytic principles of CA and the findings they uncover.
2. Throughout this study, the term “chat room” is used to refer to this specific type of chat room – online, multiparty, voice-based Skypecasts.

3. This differs from the “conference call” feature Skype offers, in which speaking participants are highlighted on the user interface.

4. If participation changes (such as by a telephone being handed to another party), this would normally be announced in advance.

5. Other forms of first turns, such as questions, can also not be responded to in the corpus, but for reasons of spatial constraints, here we only consider instances of first summons, in the format of “hello” or “hi,” being not responded to.

6. The names shown on the transcript are pseudonyms. While exactly what was originally said is not the same in the transcript, how it was said is represented as closely as possible, in terms of the number of syllables in the name and the intonation used in its pronunciation.

References


Appendix: CA Transcription Conventions

[ ] Overlapping utterances (beginning [ ) and ( end ] )
= Contiguous utterances, or continuation of the same turn by the same speaker even though the turn is separated in the transcript
(0.2) The tenths of a second between utterances
(.) A micro-pause (1 tenth of a second or less)
: Sound extension of a word (more colons demonstrate longer stretches)
. Fall in tone (not necessarily the end of a sentence)
, Continuing intonation (not necessarily between clauses)
- An abrupt stop in articulation
? Rising inflection (not necessarily a question)
__ Emphasized word or sound
↑↓ Rising or falling intonation
° ° Talk that is quieter than surrounding talk
hhh Audible aspirations
.hh Audible inhalations
(hh) Laughter within a word
> > Talk that is spoken faster than surrounding talk
< < Talk that is spoken slower than surrounding talk
(( )) Analyst’s notes
( ) Approximations of what is heard
$ $ Talk uttered in a ‘smile’ voice

(Modified from Atkinson and Heritage, 1984)

Biographical Notes

Adam Brandt [adambrandt[at]me[dot]com] is a postdoctoral research fellow at Kansai University, Japan. He is interested in researching naturally-occurring social-interactional conduct and language in use, particularly in contexts and settings which pertain to international encounters, second language use, and the use of new technologies.

Christopher Jenks [cjjenks[at]cityu[dot]edu[dot]hk] is an associate professor in the Department of English, City University of Hong Kong. His main research approach is microanalysis (e.g., conversation analysis and interactional sociolinguistics). His research deals primarily with computer-mediated communication, intercultural communication, English as a lingua franca, and second language acquisition.