Temporal Properties of Turn-Taking and Turn-Packaging in Synchronous Computer-Mediated Communication

Claude G. Čech
Department of Psychology and
Institute of Cognitive Science
University of Louisiana at Lafayette
cech@louisiana.edu

Sherri L. Condon
The MITRE Corporation and
Institute of Cognitive Science
University of Louisiana at Lafayette
scondon@mitre.org

Abstract

Turn structure and timing are examined in a variety of quasi-synchronous computer-mediated interfaces. The message window size, presence of scrolling, a single message window vs. message windows for each participant, and message persistence were systematically varied for pairs of interlocutors engaged in the same decision-making task. Participants produced more total words and more turns in conditions with larger windows and in those with scrolling, while separate windows conditioned even larger increases on these measures. Turn sizes were smaller in the latter conditions and response times were faster. In the persistent separate-window conditions, messages from the partner intervened before participants completed responses in over half of the messages.

1. Introduction

The characteristics of ordinary face-to-face conversation reflect adaptations to the constraints of the communicative environment: acoustic signals decay rapidly, and if more than one message is transmitted at the same time, there is a low signal to noise ratio. Cognitively, face-to-face interaction demands that participants simultaneously decode incoming messages and encode replies to those messages, which requires maintaining the coherence of the discourse [1] and constant grounding of the talk at several levels [2]. At the same time, participants must identify appropriate positions in the sequence of signals for transmission of their messages.

All of these physical, cognitive, and linguistic factors may have contributed to the emergence of the turn-taking conventions that participants adopt in conversation [3] and to the typically short, rapid exchange of turn packages that is observed in face-to-face interaction [4]. Computer-mediated interaction makes it possible to observe the influence of some of these factors because researchers can systematically vary features of the communication environment. These features can include synchronicity [5,6,7], message persistence [8], number of participants [9,10], and many other characteristics of the computer interface [11,12].

For the studies reported here, four features of the communication environment are systematically varied, though all conditions involve dyads engaged in the same decision-making task. We compare the size of the window in which a message may appear or be composed, whether separate windows are used for incoming and outgoing messages, whether messages in a window may be scrolled, and whether a partner's message is available while typing a reply. In a previous study [13], participants produced significantly shorter turns when the window in which they produced and received messages held 4 lines compared to 10 or 18 lines. For the current studies, we again vary window size, this time looking at 4-line and 10-line windows, and unlike the earlier study, we allow messages to extend beyond the current window size. We also compare conditions in which the interface does or does not permit scrolling of messages within the windows. These variables allow us to examine whether participants produce shorter messages in smaller windows because the window does not hold longer messages or because they would be unable to view all lines of longer messages in the smaller window. Thus we specifically vary these potential constraints to see whether removal of one or both will result in longer messages in the smaller windows.

Our previous studies [14] used a single window in which each new message replaced the previous one, so that the partner's message disappeared as soon as the participant began to type a reply. This design reduces the permanence of the messages and makes the environment more similar to face-to-face conversation, but it is atypical for most computer mediated environments. Accordingly, we included conditions in which messages from the partner appear in a separate window. Using separate windows for each participant's messages also allowed us to vary message persistence by including a condition in which the partner's message appears in a separate window, but still disappears when the participant begins to type a reply.

For the different conditions that result from varying these features, we compared measures of messages such as the average message size and the average number of messages per interaction. Some of these comparisons have been reported previously [12,15]. However, we also examine temporal data for the first time. Each message is recorded with the time that the participant began to type and the time that the message was sent, which provides a rich representation of the message exchanges and additional insight into participant's communicative strategies.

2. Communication Environment Features

Synchronicity is a factor that has been discussed by several researchers [9,16] The terms *synchronous* and *asynchronous* are introduced in [5] to contrast computer-mediated interaction in environments like chat and instant messaging with interaction via e-mail or bulletin boards. In the former, messages receive immediate replies, while the latter involve unpredictably long delays between contributions to the interaction. The term *quasi-synchronous* is introduced in [6] for synchronous transmission in which messages are composed before transmission as opposed to transmission of each character as soon as it is typed, as in the Unix Talk function. Comparisons of second language students discussing readings in a chat environment and a bulletin board environment show that they produced longer and more syntactically complex messages in the asynchronous environment, but a greater variety of discourse functions in the synchronous environment [7].

Dramatic differences between quasi-synchronous and asynchronous decision-making interactions are reported in [17]. The shorter messages in the synchronous conditions did not have a characteristic discourse structure, while the long e-mail messages had a characteristic structure that resembles the patterns observed in [18] for posts to a listserve. Messages began with strategies that managed the discursive and interpersonal work of the interaction: openings, references to previous messages, and pleasantries. They contained a body in which the decision-making work was accomplished, and they ended with strategies that managed the discursive and interpersonal work: closings, references to future messages *(I'll write again soon),* and pleasantries, especially about future events, e.g. *I hope you have a great Easter vacation*. In the studies reported here, all interactions are quasi-synchronous, and they lack the predictable structure observed in asynchronous messages.

We are not aware of previous studies that systematically vary features of the message windows. It has been suggested that one reason turns in chat interactions tend to be short is the fact that "many interfaces provide a relatively shallow window in which to compose, and so composing a long message means that the first part of one's message will scroll out of sight before one has finished" [19]. This is the kind of reasoning that motivated us to vary window size and scrolling in the studies reported here. We have suggested that cognitive demands play a role in the preference for short turns in face-to-face interaction, but computer-mediated messages do not decay instantaneously, and it is not usually necessary or even possible to monitor partners' turns for appropriate positions to transmit one's message. Consequently, some of the features favoring short turns are not present in computer-mediated interaction, though other factors that might constrain turn size, such as the size of the composition window, may come into play.

At the same time, there are many other reasons why packaging turns into short messages would be a successful strategy for synchronous interaction: working memory constraints may make it easier to process new information in smaller chunks, for example. Similarly, it is claimed that speakers must meet a grounding criterion at several different levels to assure successful communication [2], and the need to monitor the success of grounding may also bias speakers towards short turns: grounding failure in a long turn may require too much backtracking to locate the site of the misunderstanding. Finally, listeners prepare responses as they attend, and there are undoubtedly attentional and other costs involved in maintaining a relevant and coherent response to turns that become lengthy (see [1] for a discussion of some of the factors and costs influencing coherence/success of a conversation). Indeed, by Grice's maxim of relevance, the response to a long turn might require extensive indexing to appropriate segments of the turn because there may be ambiguity regarding which segment the response is relevant to.

But if there are specific features that predispose participants towards short turns, there are also identifiable features that potentially predispose them towards long turns. For example, one factor that favors longer turns is message persistence, which interacts with synchronicity. Asynchronous interaction requires messages that can persist through the delays between transmission and receipt of messages, while synchronous face-to-face interaction succeeds with minimal persistence. By comparison, the persistence of messages in most synchronous computer-mediated interactions makes them more similar to asynchronous communication. Even the single-window environment in our studies allows participants to take as much time as needed to read their partners' messages and plan their replies before they begin to type and the messages disappear. Therefore, message persistence relieves some of the demands for simultaneous processing that are present in face-to-face interaction and makes production of longer messages more feasible.

Many factors might influence the strategies that participants adopt for encoding and packaging their contributions to the interaction. Therefore, it may not be possible to isolate simple effects, even in the controlled environments of our studies, in which participants communicate about identical topics using identical interfaces. For example, people may have evolved such efficient processing and packaging strategies in the demanding face-to-face environment that they function more effectively in communication environments in which they can take advantage of those finely tuned, ingrained practices. Alternatively, some features of the communication environment may allow participants to take greater advantage of highly effective reading strategies such as skimming. By examining the rate of interaction, we can formulate inferences about processing speed, and we can also use the time tags to make observations about turn management. Consequently, we examine temporal information about the interactions to provide some additional insight into the complex interplay between the communication environment and the strategies that interlocutors adopt to manage their interaction.

3. Methods

Students from the University of Louisiana at Lafayette's Psychology Department Subject Pool were randomly assigned to conditions in the experiments reported here. All participants received identical instructions for the decision-making task that they performed over computers connected together in a LAN. Some participants were in separate rooms, while others were in one large common room, but screens prevented them from seeing other participants' monitors and from visually identifying their partners.

In Experiment 1, 4 groups of 20 dyads planned the MTV Music Video Awards by deciding who will host the show, the specific categories of music video awards to present, the nominees, the winner and the entertainment between award announcements. Two groups had a message composition area consisting of 4 lines (of 80 characters), and two groups had a 10-line composition area. All groups could type a message much longer than the composition area, but could only view portions of the message that fit in the composition window. However, one group could scroll the text in the window to view portions of the message that did not fit on the viewable 4 or 10 lines, whereas scrolling was disabled for the other group.

Participants composed their messages before pressing a dedicated SEND key. To create conditions that resemble the impermanence of messages in face-to-face environments, messages are kept minimally persistent by allowing the message composition area to hold only the current message. Therefore, when the participant receives a message, it replaces the text in the composition area, and it disappears as soon as the participant begins to type a reply.

The minimally persistent single-window condition increases demands on turn management because the consequences of unsuccessful management can compromise understanding, just as talking at the same time can compromise the success of face-to-face interaction. In the conditions for Experiment 1, if the partner sends a message while the participant is typing, it replaces the participant's message, producing the kind of frustration expressed by the participant who wrote, "please don't type while I'm typing because it messes up my message." In previous studies this problem was compounded by the fact that participants became impatient while waiting for their partners to type replies, sending messages such as "are you there?" to insure that the partner was still connected and engaged. To minimize these problems, the interface includes a message status area that informs participants when their partners are composing messages. This feature was used in all interfaces for all experiments.

In Experiment 2, messages were more persistent because each participant's compositions appear in a separate, scrollable 4- or 10-line window. Consequently, incoming messages do not replace text in the participant's composition area, and they are available for consultation during composition of a reply. However, each window contains only the most recently produced message: participants cannot scroll back to previous messages and replies. The interface in Experiment 3 is the same as in Experiment 2, except that messages are not scrollable, and persistence is minimized by deleting the partner's message as soon as the participant begins to produce a reply. Therefore, unlike Experiment 1 and like Experiment 2, participant's compositions are not replaced by intervening messages that the partner transmits. However, like Experiment 1 and unlike Experiment 2, the partner's message is not available for consultation after a participant begins to reply.

4. Results: Turn Structure

Table 1 presents the packaging features of the interactions in the 8 conditions examined across the 3 experiments. In order to characterize these features, it is necessary to clarify some terminology. The term *turn* usually refers to a stretch of speech bounded by speech from different speakers: turns are identified by changes of speakers, though turn boundaries may be blurred by simultaneous speech and backchannel contributions in face-to-face conversation. In the data reported here, participants may send more than one message before the partner replies. Therefore, when it is important to distinguish these units, we reserve the term *broadcast* to refer to a single message transmitted to the partner and use *turn* to identify one or more broadcasts bounded by a change of sender. The term *message* is used more loosely to refer to participants' contributions to the interaction.

Beginning with the conditions in the single-message window of Experiment 1, we can observe that participants packaged their contributions into similar sizes of broadcasts and turns, measured in words, in all four conditions. The larger window sizes of the composition area did not result in larger messages, nor did the ability to scroll portions of messages that exceeded the window size. These results were unexpected because a previous study had found that both turn and broadcast sizes increased when the composition area was increased. Also, participants did seem to be aware of the advantages of scrolling, as indicated by the message in (1), where the text prior to the last asterisk identifies the participant and the initiating and terminating times, and the text is presented uncorrected, as it was received.

(1) P1:04:41:17p*04:44:12p*r&b male Brian Mcknight R&B female Alicia Keys. I don't remember the rest because i did not know that once i scrolled down the page i could not go back up, so try to type a little at a time. Sorry

We speculate that the expectations and experience reflected in (1) may have resulted in participants' preference for packaging their contributions into relatively short broadcasts and turns. When the previous studies were conducted seven years earlier, students had not grown up with Instant Messaging and wireless text messaging, where short messages are clearly the norm. Even in the earlier studies, participants who packaged their contributions in the smallest units were distributed evenly across the three sizes of message areas, suggesting that the adoption of a short-message packaging strategy was independent of message-area size. Only interactions with the largest packaging units were distributed disproportionately in the conditions with larger message-area sizes. Of course, as the long message in (1) illustrates, participants used long turns, though not for a majority of messages.

However, the difference in window sizes did lead to a significant difference in packaging strategies in the current data if message size is measured in lines rather than words. Unlike the previous studies in which the "Enter" key was also the "Send" key, participants could use the "Enter" key to format their messages as in (2).

(2) P2: 11:22:48 a*11:24:37 a* What about the rest of
the categories we need at least seven categories
Like Best all around performer
Best video
Best group of the year
We have to make the award show interesting.
How about you do one thing and I do the other

Participants frequently used the spatial orientation on the page for list structures or to separate distinct functions, as in the three last lines of (2). Broadcasts in the 10-line condition were significantly longer than those in the 4-line condition when measured in lines (1.26 versus 1.147), $F\ (1,76) = 4.04$, $p < .05$. If we examine the longest broadcasts that participants composed measured in lines, the average maximum broadcast in the 10-line condition was 5.72 lines compared to 3.35 lines for the 4-line condition, $F\ (1,76) = 9.50$, $p < .005$. Therefore, the data suggest that the larger composition area facilitates participants' exploitation of the graphic resources available to them in the computer-mediated environment.

Furthermore, both the larger message-area size and the scrolling function appear to have facilitated the interaction by encouraging participants to contribute more. Participants in the scrollable conditions tended to transmit more in terms of broadcasts and turns, although the results for turns did not quite reach conventional levels of significance: for broadcasts, $F\ (1,76) = 4.68$, $p < .05$; for turns, $F\ (1,76) = 3.56$, $p = .063$. Similarly, participants in the 10-line conditions tended to transmit more: for broadcasts, $F\ (1,76) = 3.81$, $p < .055$; for turns, $F\ (1,76) = 6.20$, $p < .02$.

Table 1: Packaging of Turns and Broadcasts in All Conditions

	Single Message Composition Area				Two Separate Message Areas			
	Scrollable		Non-scrollable		Persistent (scrollable)		Non-Persistent (non-scrollable)	
	4-line	10-line	4-line	10-line	4-line	10-line	4-line	10-line
File size/words	523.6	700.6	453.4	535.6	718.6	788.8	754.1	842.8
No. of broadcasts	42	66	40	41	80	97	88	87
No. of turns	31	51	30	34	58	66	62	60
Broadcast size in words	14.4	13.6	13.0	15.6	10	9.4	10.3	11.8
Turn size in words	18.2	17.5	16.4	18.4	14	13.5	13.8	16.5
Max Broadcast Size	46.6	49.4	47.4	55.4	38.3	37.8	41.8	46.8
Broadcasts/Turns	1.4	1.3	1.3	1.2	1.4	1.5	1.4	1.5

There is a striking difference in Table 1 which suggests that these effects are predominantly due to the 10-line scrollable condition. However, the p-levels for the interactions in both cases failed to reach the conventional .05 level (both were between .07 and .09).

Similar results are obtained if we examine the total words and characters produced in the interactions. Whether measured by total number of words transmitted or by total number of (non-space) ascii characters, the 10-line and scrollable conditions led to more transmitted language: in words, F (1,76) = 8.83, $p < .005$ for the variable of window size, and F (1,76) = 7.27, $p < .01$ for scrollability; in ascii characters, these respective F-values were 7.53 ($p < .01$) and 6.34 ($p < .02$). In these analyses, the interaction of window size with scrollability did not approach significance.

If increases in the amount of communication indicate that the interface has facilitated interaction, then the design with two message-areas has a strong positive impact in the data from Experiments 2 and 3. For convenience, we will refer to this type of interface as the 2-channel design, foregrounding the contrast between the two areas available in the computer-mediated interface and the single auditory channel available in face-to-face environments. To enable comparison of single- and two-channel conditions, we first ran analyses of variance comparing Experiments 2 and 3 on the data of Table 1. These analyses displayed no significant main effects or interactions.

Based on the comparability of Experiments 2 and 3, we then ran an omnibus analysis that involved the factors of channel number, window size, and scrollability. As Table 1 illustrates, the numbers of broadcasts, turns, and words in the 2-channel environments dramatically increase compared to the single-channel environment. The 2-channel conditions average 88.2 broadcasts and 61.6 turns versus 47 broadcasts and 36.6 turns for the single channel design: for broadcasts, F (1,152) = 47.2, $p < .0001$; for turns, F (1,152) = 38.84, $p < .0001$. Similarly, participants in the 2-channel interface averaged about 223 more words per interaction, F (1,152) = 30.15, $p < .0001$ and they also sent more (non-space) ascii characters, F (1,152) = 29.88, $p < .0001$.

The larger interactions in the 2-channel interface were packaged into smaller transmission units. The average size of a broadcast in the 2-channel conditions was slightly over 10 words, compared to the 14.2 words of the one-way procedure, F (1,152) = 12.41, $p < .001$. Similarly, the average turn size was smaller in the 2-channel conditions (14.5 vs. 17.6 words), F(1,152) = 6.26, p < .02. As is evident from the ratio of broadcasts to turns in the last row of Table 1, participants in the 2-channel interface were more likely to construct turns consisting of several short serial broadcasts. Moreover, they were more likely than participants in the single channel condition to write single-line messages, F (1,152) = 15.39, $p < .0002$; these constituted 85.6 % of the messages sent, compared to 77.6 % in the single-channel condition.

Therefore, the variables of window size and scrollability had minimal effects on the packaging strategies that participants adopted in the 2-channel interactions. Because participants favor packaging their contributions in smaller transmission units in the 2-channel conditions, it is not surprising that increasing the composition area fails to influence message sizes. Moreover, the two independent variables that were affected by the increase in message-area size in the single-channel conditions were the average file size and the average numbers of transmission units (broadcasts and turns). Yet the 2-channel condition influences these variables to reach values that are, with one exception, considerably higher than any values obtained in the single-channel condition. Therefore, it may be that using 2 channels raises these values to maximal levels that resist any additional enhancements.

Similar reasoning can explain why we do not observe packaging effects from reducing persistence in the 2-channel conditions. It might be expected that interaction would be facilitated when the partner's message can be consulted while producing a reply, yet we do not observe increased interaction. Possibly, messages are so minimal in the 2-channel condition that there is no need to consult them while producing a reply: they are easy to remember. After all, people are routinely successful in managing non-persistent messages in face-to-face interaction. Nevertheless, other factors appear to be involved, too, as we examine temporal differences among the conditions.

6. Temporal Properties

Table 2 presents information obtained by examining the temporal information that accompanies each broadcast. In particular, we were interested in the interval between the receipt of a message and the start of the reply to the message, reasoning that participants should take longer to read messages and plan their replies if messages disappear as soon as they begin to type a response. Therefore, it was anticipated that the time to begin a reply would be longer in the single-channel conditions and the non-persistent conditions in the 2-channel environment.

To measure the lag between receipt of a message and start of the reply to that message, the start time is subtracted from the sent time of the previous message. This measure is complicated by two factors. First, the subtraction is performed only when the previous message is from a different participant. Second, sometimes participants received messages from the partner while composing their messages. In these cases, a negative value is obtained if the time that the participant begins a message is subtracted from the time that the partner sends the most recent message. Consequently, we tracked these values separately, using the positive values in the first row of Table 2 labeled as "Time to Begin Reply" and the negative values in the last row of Table 2 labeled as "Time to Receive Intervening."

Two additional caveats ought to be mentioned. First, because we were not interested in millisecond accuracy, these times were based on the system clock for a Visual Basic program running in Windows mode. The background multitasking in Windows mode will add a variable delay to the real times, although at the coarse timing level with which we are concerned, that delay will be inconsequential. Second, due to traffic on our network, there was often a significant delay between the time a participant sent a message, and the time that the partner received it. Each computer tracked (1) the time at which it received a signal that the partner on another computer started composing a message, (2) the time at which that message was finally received on the host computer, (3) the time at which the participant on the host computer started composing a message, and (4) the time at which that message was sent. Thus, the reported times and results are always conditioned on each participant's host computer.

Beginning with the positive values, the "Time to Begin Reply" measured in seconds is unaffected by persistence in the 2-channel environment, $F(1,156) = 1.5$, $p > .20$, and exhibits a non-significant trend towards a longer lag time in the 10-window condition, $F(1,156) = 2.25$, $p < .15$. But the table shows that the lag time is clearly greater in the single-channel conditions compared to the persistent 2-channel conditions, as predicted: $F(1,312) = 58.02$, $p < .0001$. Moreover, a significant interaction between number of channels and window size, $F(1,312) = 5.41$, $p < .02$, reflects a window size effect in the single-channel condition: a tendency for longer lag times in the 4-line window. Of course, these values do not take into account the fact that broadcasts and turns were longer in the single-channel conditions. As an approximate correction for this fact, the next two rows of Table 2 present the average times to begin divided by the average number of words in the broadcasts and turns. When the times are adjusted in this way, the relations remain the same: the single-channel lags are greater and there is no difference between persistent and non-persistent conditions in the 2-channel environments.

We can obtain an approximation of reading and planning rates by inverting the ratios. If the number of words in the average broadcast or turn is divided by the average number of seconds that participants took to reply, we obtain the values in the rows labeled "Broadcast Size/Time to Begin" and "Turn Size/Time to Begin" in Table 2. This expresses the ratios as words per second, which makes it possible to compare them to measures of reading and speaking rates. Reading rates from 200 to 400 words per minute (3.3 words per second) are reported in [20]. Speaking rates of about 180 words per minute or 3 words per second are reported in [21]. Of course, participants are planning in addition to reading, and these ratios are only rough approximations of the actual rates.

We adopt the term *intervening message* for messages that are received by participants while they are composing messages. It seems advisable to avoid the assumption that these are interruptions in the ordinary sense, especially considering the occasional delays in routing messages on the network. In many cases they appear more like instances in which speakers begin to talk at the same time, and one

Table 2: Timing of Turns and Broadcasts in All Conditions

	Single Message Composition Area				Two Separate Message Areas			
	Scrollable		Non-scrollable		Persistent (scrollable)		Non-Persistent (non-scrollable)	
	4-line	10-line	4-line	10-line	4-line	10-line	4-line	10-line
Time to Begin Reply*	16.54	12.46	13.78	14.00	8.44	9.40	9.21	10.31
Time to Begin/Broadcast Size	1.15	.92	1.06	.9	.84	1.0	.89	.87
Time to Begin/Turn Size	.91	.71	.84	.76	.60	.70	.67	.62
Broadcast Size/Time to Begin	.85	1.09	.94	1.11	1.18	1.00	1.12	1.14
Turn Size/Time to Begin	1.1	1.4	1.2	1.3	1.7	1.4	1.5	1.6
Percent with Intervening	18.3	19.8	14.1	12.2	49.4	54.8	38.0	43.1
Time to Receive Intervening	33.40	30.74	32.08	43.11	37.26	31.69	26.24	35.02

*turns without intervening messages only

speaker then stops while the other continues. The sequence in (3) provides several examples.

(3) a. P1: 04:11:33 p*04:11:50 p*i was thinking of makeing fun of the musci videos
b. P2: 04:11:55 p*04:12:04 p*ok...which ones
→ c. P1: 04:11:51 p*04:12:29 p*like Nelly's Hot N' Here
d. P2: 04:12:43 p*04:12:45 p*sounds good
e. P2: 04:12:55 p*04:13:05 p*make sure he has the bandaid on the cheek though like nelly..haha
→ f. P1: 04:12:54 p*04:13:51 p*yeah. that should be good. Ok let's try for the categories next. What should be the first one?
g. P2: 04:14:05 p*04:14:23 p*how about "best pop video”
h. P2: 04:14:27 p*04:14:33 p*who should present it
→ i. P1: 04:14:27 p*04:14:53 p*sounds good to me. I think Lenny Kravitz should present it
j. P2: 04:15:02 p*04:15:05 p*k...lenny it is

In (3c) P1 begins a second broadcast one second after sending (3a). Four seconds later, P2 begins to compose a reply to (3a), and we cannot know whether P2 was aware that P1 had begun to compose the second broadcast, even though the message status area on P2's screen would indicate that there is an incoming message. In these decision-making interactions, the default continuation for a suggestion like (3a) is an agreement like (3b) [14,22]. Therefore, P2's agreeing turn immediately after P1's suggestion is the expected, unmarked behavior for this state of the interaction, and P2 may have been so intent on processing (3a) and producing (3b) that she was genuinely not aware that P1 was producing a second broadcast.

Similarly, it is not possible to be certain that P1 was aware of P2's incoming message, though clearly P1 received (3b) before sending (3c), perhaps even changing the message that she originally planned in order to respond to (3b). When P1 finally completes (3c) 25 seconds after P2 sent (3b) and 38 seconds after starting (3c), the content does seem to be responding to (3b), though this may be a case of phantom responsiveness, as described in [23].

The next intervening message occurs after (3d). The default continuation for an agreement like (3d) is to move on to the next decision, as P1 does in (3f). However, within a second of P1's starting (3f), P2 produces the second broadcast in (3e), and it is not clear whether (3f) is responding to (3e). The third intervening message occurs after P2 produces a suggestion in (3g). At the same time that P1 begins a response to (3g), P2 starts a second broadcast that solicits a suggestion in (3h). In this case, it seems clear that P1 responds to both (3g) and (3h) when she completes her response in (3i).

The exchanges in (3) suggest that intervening messages primarily occur when participants deviate from a norm in which each broadcast is a turn. Instead, their turns consist of more than one broadcast produced in sequence. The broadcasts can be viewed as analogous to turn construction units [3], each of which is potentially, but not necessarily, a single turn. In ordinary conversation, participants are adept at anticipating when the end of a turn construction unit is also the end of a turn, and they typically do this by integrating abstract linguistic information with prosodic and visual cues such as gaze. If they miscalculate and speak simultaneously, then they often adopt the strategy observed in (3), in which one participant pauses to let the other continue.

Unlike face-to-face interaction, computer-mediated interaction does not include prosodic information or visual cues obtained from physical co-presence. Moreover, the packaging of a turn construction unit in a broadcast is likely to be interpreted as a cue that the turn has ended, especially in an environment in which turn-taking norms have not been conventionalized. In at least one machine-mediated communication environment, telephone devices for the deaf, signaling ends of turns emerged as such an important management function that a conventional marker *ga* "go ahead" was adopted. Consequently, we might expect that participants would avoid multi-broadcast turns and intervening messages in order to minimize the potential turn management problems associated with these strategies. Instead, Table 2 shows that intervening messages are frequent in the 2-channel environment. In the persistent conditions, about half of the turn changes involve intervening messages. In contrast, intervening messages are involved in less than 15-20% of the turn changes in the single-channel environment. Similarly, participants produce more multi-broadcast turns in the 2-channel conditions, as evidenced by the higher ratios of broadcasts to turns in the last row of Table 1.

In the single-channel environment, intervening messages were less likely to occur in the non-scrollable conditions $F(1,156) = 5.08$, $p < .05$, and in the 2-channel environment they were less likely to occur in the non-persistent conditions, which are also non-scrollable $F(1,156) = 11.97$, $p < .001$. An ominbus anova including channels as a factor confirms the significant difference between the single-channel and 2-channel conditions $F(1,312) = 203.84$, $p < .0001$ and the significant effect of scrollability/persistence $F = 16.93$, $p < .0003$. In all of these cases, higher proportions of intervening messages can be viewed as the result of increased persistence. Messages are more persistent in the scrollable conditions because participants can scroll to text that would otherwise be inaccessible outside of the message area, and they are more persistent in the 2-channel conditions because the single-channel environment displays only one message at a time.

The 2-channel design clearly supports different strategies for turn management than the single-channel environment. The 2-channel design seems to make turn management easier, and this result is not achieved by a device like the message-status area that was added to facilitate turn management in the environments designed for these studies. Instead, turn management is easier because turns need not be as carefully managed in 2-channel environments compared to single-channel environments: the consequences of simultaneous transmission are less catastrophic. Consequently, participants are not constrained by rigid turn-taking requirements, and they are free to experiment with alternative turn management strategies.

In computer-mediated environments where messages can be persistent, the consequences of simultaneous production do not compromise the success of the transmission, and participants can potentially produce their messages simultaneously throughout the interaction. Nevertheless, participants do not engage in large proportions of simultaneous transmission, and there are significant factors that motivate a sequential organization of the interactions. The nature of interaction itself requires that participants respond to each other's contributions, which necessitates a sequence of events. The times recorded in the final row of Table 2 suggest that in most instances of simultaneous production of messages, participants behave as they do in (3), with participants delaying the completion of messages in order to decode and respond to partners' intervening messages.

A major motivation for preserving the sequential organization of decision-making interactions is the decision routine that is adopted to perform the task. It is well-documented that participants rely on a decision routine in which the goal of each decision is established as in (3b,f,h), a suggestion satisfying the goal is produced by one of the participants as in (3a,c,g,i) and, in the familiar structure of an adjacency pair, an agreement is produced by the other participant, as in (3b,d,i,j) [14,22]. The sequencing of this routine follows the logic of the decision process, and the conventionalization of this process into a discourse routine provides a powerful tool for managing the interaction. The routine activates a frame for recognizing the discourse functions that are linked in the routine and for understanding the consequences of those functions in the decision process. This knowledge becomes part of the common ground and allows participants to perform the routine functions with relatively low expenditures of linguistic and cognitive resources, such as "ok" in (3b).

Routines operate by activating expectations about default continuations, and default continuations are, by default, adjacent. Therefore, adjacency pairs are special cases of routines: the default continuation of the first pair-part of an adjacency pair is the second pair-part, just as the default continuation for a suggestion is an adjacent agreement. Not only is adjacency an important resource that turn management strategies might seek to preserve, but also expectations about adjacency can interact with turn management. For example, if a turn construction unit consists of the first pair-part of an adjacency pair, such as a question, it should be more likely that the speaker will not continue the turn with another turn construction unit because the default is for the partner to produce the second-pair part. These examples illustrate that turn management is not simply an exercise in transmission logistics. Turn management is a significant tool in performing the work of the interaction, and it is influenced by factors that are central to the success of the communication.

7. Conclusions

The results presented here demonstrate that features of the communication environment influence the strategies that participants in interaction adopt to package their contributions into transmission units. The analyses also reflect the complexity of factors that are coordinated by turn management strategies and some of the trade-offs that designers of communication environments must consider. The cognitive and linguistic mechanisms that have evolved for face-to-face interaction have been refined into highly effective practices for successful communication in an environment with extremely demanding constraints on processing. Speakers must simultaneously decode their conversational partner's contributions, plan their replies, and anticipate turn boundaries so that they can transmit their contributions with split-second timing. The organization of ordinary conversation in short, rapidly-exchanged turns reflects these demands, which arise from the impermanence of speech signals, the constraints of the single auditory channel, and the need to coordinate contributions to a joint project [2].

Computer-mediated environments can relax some of the processing demands that interlocutors must satisfy. When transmission is not confined to a single channel, careful synchronization of turns is not essential. However, this relaxation of constraints on turn-taking comes with a tradeoff because turn timing can no longer be used as a resource for interactional goals. For example, backchannel acknowledgments cannot be produced with the perfect, non-disruptive timing that occurs in face-to-face interaction. Functions associated with interruption, overlap, or completing the partner's utterance must be accomplished by other means, and it is likely that people will devise new ways of achieving their goals by exploiting the new resources available in computer-mediated environments. In the studies reported here, participants in the 2-channel environments use multi-broadcast turns and intervening turns with a regularity that suggests they were finding some useful functions for those turn management strategies.

Another trade-off in turn management arises from the need to coordinate contributions in joint activity and the significance of adjacency for organizing both the form and content of interaction. Turn-taking and turn sequencing cannot become so relaxed that it prevents participants from achieving their goals or exploiting effective management practices such as the decision-making routine. One reason why multi-broadcast turns and intervening messages may have been tolerable for the participants in these studies is that there were only 2 interlocutors in the interactions. It has been observed that the resource of adjacency is lost in chat interaction with large numbers of participants, making coherence difficult to maintain [24]. In another study with only 3 participants, video records are used to show that messages might be interpreted as responding to previous messages when in fact they are not (phantom responsiveness), which can result in what the authors call *phantom adjacency pairs* [23]. In one example, when the second pair-part of a phantom adjacency pair cannot be interpreted as a relevant reply to the first pair-part, the questioner responds "What?!!!" This example illustrates that simultaneous transmission can be catastrophic when it violates the adjacency expectations of discourse routines.

It seems likely that in Experiments 2 and 3, the 2-channel interface with only two participants made it easier for participants to monitor the sequence and timing of intervening messages so that adjacency is not compromised. According to this reasoning, the trade-off between simultaneity and adjacency might be resolved by an interface like the one proposed in [11] in which each participant's messages occupy a separate track on the y-axis of the screen, while the x-axis location reflects relative start times. Clearly, future studies of interactions with larger numbers of participants are needed to examine the effects of additional interlocutors on the variables we have manipulated.

Perhaps the most significant trade-off observed in these studies has important implications for designers of communication systems. It is understandable that researchers might view face-to-face interaction as a gold standard for successful communication, and it is likely that the processing mechanisms which have evolved for face-to-face interaction are so efficient that any communication system which allows participants to take advantage of these abilities will achieve enhanced performances. The data reported for the 2-channel environments provide some support for this reasoning because those were the conditions in which participants packaged their contributions into messages that most closely approximate the average 8.6 words per turn produced by participants who performed the same decision-making task in face-to-face interaction [25]. These participants not only produced more messages, but they processed them more efficiently. Consequently, it appears that the 2-channel design allowed participants to transmit their messages in a way that taps into highly efficient processing mechanisms for decoding and planning contributions.

Though participants performed better in the 2-channel conditions by measures of processing speed, file size, and numbers of transmission units, there is still one strategy that emerged in the single-channel conditions, but seems to have been discouraged in the 2-channel environment. The use of graphic resources that occurs when participants break their messages into separate lines is necessarily reduced in the brief broadcasts of the 2-channel conditions. It appears that when participants are employing strategies and processing mechanisms that are more similar to face-to-face interaction, they are also less likely to explore alternatives that exploit the rich resources available in the text-based environment. Consequently, by encouraging participants to operate more like they do in face-to-face contexts, designers might be discouraging them from developing practices that realize the full potential of persistent conversation.

10. References

[1] Clark, H. and Brennan, S. 1991. Grounding in communication. In L. Resnick, R. Levine, and S. Teasley (Eds.), *Perspectives on Socially Shared Cognition,* APA Press.

[2] Clark, H. 1996. *Using Language*. Cambridge: Cambridge University Press.

[3] Sacks, H., Schegloff, E., and Jefferson, G. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language, 50*, 696-735.

[4] Schegloff, E. 1982. Discourse as an interactional achievement: Some uses of *uh huh* and other things that come between sentences. In D. Tannen (Ed.), *Analyzing discourse: Text and talk.* Washington, D.C.: Georgetown University Press.

[5] Kiesler, S., Zubrow, D., Moses, A., and Geller, V. 1985. Affect in computer-mediated communication: An experiment in synchronous terminal-to-terminal discussion. *Human-Computer Interaction, 1*, 77-104.

[6] Garcia, A. and 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.

[7] Sotillo, S. 2000. Discourse functions and syntactic complexity in synchronous and asynchronous communication. *Language Learning and Technology,* 4, 82-119. http://llt.msu.edu/vol4num1/sotillo/default.html

[8] Erickson, T. Smith, D. N., Kellogg, W. A., Laff, M. R., Richards, J. T., and Bradner, E. 1999. Socially Translucent Systems: Social Proxies, Persistent Conversation,and the Design of 'Babble.' In *Human Factors in Computing Systems: The Proceedings of CHI '99.* ACM Press.

[9] Cherney, L. 1999. *Conversation and community: Chat in a virtual world.* Cambridge: Cambridge University Press.

[10] Gray, C. 1999. Does group size in electronic mail affect communication efficiency and flaming? M.S. thesis, Dept. Psychology, U. Louisiana at Lafayette, Lafayette, LA.

[11] Shankar, T., VanKleek, M., Vicente, A., and Smith, B. 2000. Fugue: A computer mediated conversational system that supports turn negotiation. *Proceedings of the Hawai'i International Conference on System Sciences.* IEEE.

[12] Čech, C. and Condon, S. 2002a. Does attentional load affect discourse management in on-line communication? *Proceedings of the 3rd International Conference on Advances in Infrastructure for E-Business, E-Science, and E-Education on the Internet*, August 2002 (CD-ROM).

[13] Čech, C. and Condon, S. 1998. Message size constraints on discourse planning in synchronous computer-mediated communication. *Behavior Research Methods, Instruments, and Computers*, 30, 255-263.

[14] Condon, S., and Čech, C. G. 1996. Functional comparison of face-to-face and computer-mediated decision-making interactions. In S. Herring (Ed.), *Computer-mediated communication: Linguistic, social, and cross-cultural perspectives.* Philadelphia: John Benjamin.

[15] Čech, C. and Condon, S. 2002b. Influencing frequency and size of exchanges in synchronous dyadic computer-mediated communication. Presented to the Society for Computers in Psychology. New Orleans, Louisiana.

[16] Baym, N. 1996. Agreements and disagreements in a computer-mediated discussion. *Research on Language and Social Interaction,* 29, 315-346.

[17] Condon, S. and Čech, C. Forthcoming. Computer mediated discourse management in three modalities. In S. Herring (Ed.), *Computer-Mediated Conversation*.

[18] Herring, S. 1996. Two variants of an electronic message schema. In S. Herring (Ed.), *Computer-mediated communication: Linguistic, social, and cross-cultural perspectives.* Philadelphia: John Benjamin.

[19] Werry, C. 1996. Linguistic and interactional features of Internet Relay Chat. In S. Herring (Ed.), *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives.* Philadelphia: John Benjamins. (p.93)

[20] Gibson, E. and Levin, H. 1975. *The Psychology of Reading.* Cambridge, Mass.: The MIT Press. (p. 539)

[21] Chafe, W. 1982. Integration and involvement in speaking, writing, and oral literature. In Deborah Tannen, (Ed.), *Spoken and Written Language: Exploring Orality and Literacy.* Norwood, New Jersey: Ablex Pub. (p. 36)

[22] Condon, S., Čech C., and Edwards, W. 1999. Measuring Conformity to Discourse Routines in Decision-making Interactions. *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics,* 238-245.

[23] Garcia, A. and Jacobs, J. 1998. The interactional organization of computer-mediated communication in the college classroom. *Qualitative Sociology,* 21(3), 299-317.

[24] Herring, S. 1999. Interactional Coherence in CMC. *Journal of Computer Mediated Communication,* 4(4) June.

[25] Condon, S. and Čech, C. 2002. If you can read this, you're a cyborg: Turn structure in virtual interaction. Presented to the Georgetown University Roundtable on Linguistics.