

## HOW PEOPLE COMMUNICATE IN A MULTITHREADED MANNER USING TEXT AND VOICE: A PILOT STUDY

KANAYO OGURA, KAZUSHI NISHIMOTO AND KOZO SUGIYAMA

*Japan Advanced Institute of Science and Technology  
{k-ogura, knishi, sugi}@jaist.ac.jp*

**Abstract.** We attempt to extend multithreaded communications from single-modal to multimodal capability to achieve dense, effective and informative communications. We have developed a multithread-oriented multimodal chat system named “ChaTELex.” ChaTELex has particular functions for specifying a related message and receiver(s) to maintain multiple topic threads. In addition, ChaTELex allows a user to send an utterance that includes text-only, voice-only or both text and voice messages. We conducted pilot experiments using ChaTELex to explore the possibilities and features of multithreaded multimodal communications. Based on the results of experiments, we confirmed that it is possible to converse in multiple-topic threads using multiple modalities. Moreover, the results suggested two advantages of using both text and voice: the text messages can clearly indicate the contents of the voice messages, and the text messages can convey intricate feelings that are difficult to express in a voice-only message.

### 1. Introduction

Oral communications involve very inefficient characteristics. Typical oral communications are held in a single-threaded manner. This is because conventional oral communications impose two rules on people. The first one is the “single-topic rule”: the participants have to share a single-topic thread and cooperatively maintain the topic. The second one is the “turn-taking rule”: only one person is allowed to speak at a time. If someone wants to say his/her opinion, he/she has to wait for a while until the current speaker finishes talking, thus making an opportunity to take the turn. As a result, there is always only one topic thread and one speaker, even in a meeting with hundreds of attendees, and thus the density of exchanged information and knowledge becomes very sparse.

Accordingly, the authors have been attempting to lessen or even remove the effect of these rules to permit oral communications in a multithreaded

manner. We investigated text-chat communications and found the mechanisms that provide multithreaded communications in the text-chat. We applied these mechanisms to voice communications. As a result, we confirmed that people could have oral communications in a multithreaded manner.

In this paper, we extend multithreaded communications from a single-modal to a multimodal capability to achieve denser, more effective and more informative communications. Text chat accepts only text messages, while our system for multithreaded voice communications (Ogura et al., 2006) accepts only voice messages. We integrated these two systems to allow multithreaded communications with multiple modalities: text and voice.

## **2. ChaTELex: A multithread-oriented multimodal chat system**

It is very difficult to distinguish simultaneous voice utterances, to memorize them, and to accurately respond to them. Due to such difficulties, we usually have to obey the single-topic rule and the turn-taking rule. In text chats, however, multithreaded conversations are often maintained for a long time. This capability is mainly provided by the “history of utterances” feature implemented in text chat systems. The history sequentially lists all messages, thus allowing users to read any utterance and respond to it at anytime (Rhyll, 2001). In addition, skilled text-chat users use several special representations to specify receivers and related topics/messages to manage complicated multiple-topic threads (Winiecki, 2003; Ogura, et. al., 2004). Applying these findings, we created a multithreaded voice communication system named “ChaTEL” and confirmed that people can orally communicate in a multithreaded manner using ChaTEL (Ogura et al., 2007).

In this work, we integrate text chat and voice chat to achieve multimodal multithreaded communications. This new system is named “ChaTELex.” Using ChaTELex, a user can send a message that consists of text only, voice only or both text and voice. Each message can be related to one of the preceding message as well as to specific receivers.

Figure 1 shows the user interface of ChaTELex. The “Rec Voice” button (1) is used to record a voice message. A text message can be input from a text box (2). It is possible to input both voice and text messages. After composing a message, it can be sent out by pushing one of four buttons (3)-(6). The “Simple Sending” button (3) is used for just sending out the composed message. The “Send to Specific Person” button (4) associates the message with (a) specified receiver(s) in the member list (16) and then sends it out. The “Reply to Sender of Selected Message” button (6) associates the message to the sender of the selected message in the message log (11) and sends it out. The “Reply to Selected Message” button (5) associates the message with a message selected in the message log (11) and sends it out.

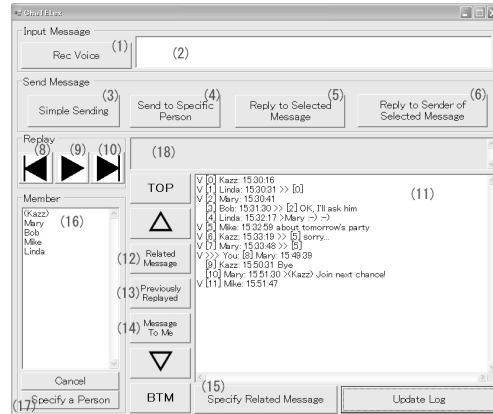


Figure 1. Interface of ChaTELex

Figure 1 (11) shows the chat log. The leftmost column in this field shows the voice flag: “V” indicates that the message includes a voice message. The next column (e.g., [4]) shows the sequence number of the message. The third column shows the handle-name of the sender. The fourth column shows the time when the message was sent. The fifth column shows, if anything, the associated receiver(s)’ name(s) and/or sequence number of the associated message. For example, “>>[2]” shows that the message is related to the second message, and “> Mary” shows that the message was sent to Mary. If the message is sent to the user, the “>>> You:” tag appears between the voice flag and sequence number (see message [8]). The final column shows the text message, if any. The entire message (except for voice data) of the selected message appears in the text window (18).

A selected message can be replayed by double-clicking it in the log (11) or by pushing button (9). Button (10) replays the next voice message, while button (8) replays the previous one. The “Related Message” button (12) replays a related message such as “>> [5].” The “Message to Me” button (14) replays a message having “>>> You” tag. The “Previously Replayed” button (13) replays the message that was listened to most recently.

### 3. Pilot Experiments on ChaTELex

#### 3.1. PROCEDURES

We conducted pilot experiments using ChaTELex. The aims of the experiments were to investigate whether multithreaded conversation could be done in the manner of ChaTEL, to observe how messages can be sent using both

voice and text, and to find particular ways of using voice and text together.

We conducted the experiments with two groups, and each group consisted of four subjects. All of the subjects were accustomed to text chat systems and had used ChaTEL. As a control system, we prepared a “baseline” system that was only equipped with the following recording/listening functions: “Rec Voice (Fig. 1 (1)),” “text input (2),” “Simple Sending (3),” “Replay the selected message (9),” and “Replay the next message (10)”; the message log (11) was also provided. The experiments were conducted in non-face-to-face situations using either ChaTELex or the baseline system. One group used the baseline system first and then ChaTELex, while the other group used them in the reverse order.

We prepared four initial topics and assigned two of the four topics to each subject; the combinations of topics for the subjects differed. In addition, we directed each subject to initiate conversation on one topic by voice and on the other topic by text. We instructed all users to converse for about 20 minutes using either the baseline system or ChaTELex.

## 3.2 RESULTS

### 3.2.1 Multiple topic threads

The average number of utterances per second was 0.07 for both systems, which was almost the same as that for ChaTEL (0.06; Ogura et al., 2007). This result shows that usability for sending messages is roughly equivalent among all three systems.

To determine whether multithreaded conversation could be carried out, we visualized the structure of topic-thread relations for each experiment’s data. Then we counted the average number of concurrent topic threads and calculated the duration of multithread situations for each number of concurrent threads. The results are shown in Tables 1 and 2.

TABLE 1. Average number of concurrent topic threads per second.

Baseline	ChaTELex
2.73	2.23

TABLE 2. Duration of multithread situations for each number of concurrent threads (sec.)

Numbers of concurrent threads	2	3	4	5~
Baseline	493	332	11	0
ChaTELex	428	450	105	377

Table 1 shows that it is indeed possible to concurrently converse on multiple topics using text and voice. Interestingly, there is a slightly higher average number of concurrent topic threads in baseline than in ChaTELex, although ChaTELex provides several functions to systematically relate the

messages and/or the receivers. In contrast, Table 2 shows that multithreaded situations are maintained longer with ChaTELex than with the baseline system for every number of concurrent threads except for the case of two concurrent threads.

ChaTELex integrated a text chat system with ChaTEL, and multiple topic threads easily arise in text chat. In addition, not only ChaTELex but also the baseline system allowed the subjects to add link information to voice utterances manually by using text. Therefore, it was easy even for the baseline to maintain two or three multiple topic threads. However, when the number of concurrent threads increased and the conversation situations became complicated, it was not adequate to simply add links to show the relations among utterances. It thus became necessary to track the links. The baseline does not provide any function to track links, while ChaTELex does (Fig. 1 (12) and (14)). We believe these functions of ChaTELex made it possible to maintain many concurrent threads for a long time.

### 3.2.2 Ways of using voice and text

First, we examined the macroscopic usage of voice and text. Table 3 shows the numbers of utterances using text and voice. “Ratio of voice” is the ratio of the number of voice utterances to the total number of utterances. “Text/Voice” means that one utterance consists of voice and text. From Table 3, we can see that voice was more frequently used in ChaTELex than in the baseline. In addition, we can see that there were more “Text/Voice” cases in the baseline than in ChaTELex.

TABLE 3. Frequencies of using voice and text. “Voice” cases include “Text/Voice” cases.

	Text	Voice	Ratio of voice	Text/Voice
Baseline	156	78	33.33%	38
ChaTELex	121	90	42.65%	10

ChaTELex’s functions of systematically relating messages should explain these results. In voice-only messages, we can understand the contents of the messages only after we listen to the messages. In the baseline, the subjects could not know the contents of the voice-only messages nor the relations between them and the other messages. Therefore, we infer that the subjects frequently added text messages to the voice messages to index the voice messages and/or their semantic relations to the other messages. In fact, we found several cases in which the subjects typed message IDs corresponding to the voice messages. On the other hand, ChaTELex’s functions (Fig. 1 (4), (5), (6)) allowed the subjects to add a message ID of the related message; this made it possible for the subjects to identify which message is related to the voice-only message without listening to it.

We further investigated the microscopic usage of voice and text. First, we analyzed how the usage of voice and text changed with time. For this purpose, we calculated a “usage score”  $S(t)$  defined as follows:

$$S(t) = \sum_{i=1}^t \delta_i,$$

where  $\delta_i$  is defined as follows:

$$\delta_i = \begin{cases} 1 & \text{if } i_{th} \text{ message is a text-only message,} \\ 0 & \text{if } i_{th} \text{ message is a voice/text message,} \\ -1 & \text{if } i_{th} \text{ message is a voice-only message.} \end{cases}$$

$S(t)$  monotonically increases if text-only messages are successively submitted, while  $S(t)$  monotonically decreases if voice-only messages are successively submitted.  $S(t)$  remains almost constant if text-only messages and voice-only messages are evenly mingled or if voice/text messages are mainly submitted. Figure 2 shows transitions of  $S(t)$  for each experiment.

As we can see from Fig. 2, the two groups have different tendencies. In the left figure, we can see that text was dominantly used in the baseline, while voice and text were evenly used in ChaTELex. In contrast, in the right figure, we can see that when text and voice were evenly used, then text was dominantly used in the baseline, while when text was dominantly used, then voice and text were evenly used in ChaTELex. We also calculated usage score  $S(t)$  for each topic thread of each experiment to examine whether the transitions of  $S(t)$  depend on the topic thread. As a result, we found that the way of using voice and text was topic-thread-dependent.

### 3.2.3 Remarkable usages of voice and text

In the log data, we found several cases of remarkable usage of text and voice, especially when text and voice are used in an utterance at the same time.

Message IDs related to voice messages are input as text messages, and this usage often occurred in the baseline. For example,

[89] User A: I cannot much drink beer, although I think it is delicious just

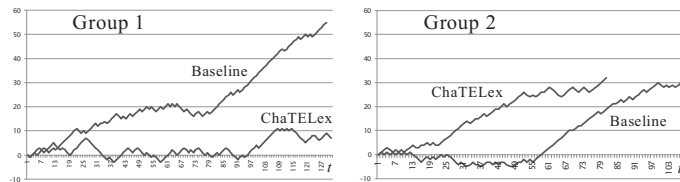


Figure 2. Transitions of usage score  $S(t)$ . Each graph shows the results of one subject group. The vertical axis corresponds to  $S(t)$  and the horizontal axis corresponds to  $t$ .

at the time I start drinking beer.

V [91] User B: >>89 (Voice message: I feel tired of drinking beer after I lift a glass to my lips...)

The 89th utterance was a text-only message, and the 91st utterance consisted of text and voice messages. The latter subject wanted to indicate to which utterance the voice utterance related in an immediately recognizable manner. This phenomenon can be considered a reason why the number of “Text/Voice” in baseline is more than that in ChaTELex (Table 3).

Text messages were also used to summarize an accompanied voice message. An example of this case is as follows.

[58] User A: I’m very hungry...

V [71] User B: Eat Fish !! >>58 (Voice message: When you are hungry, I recommend eating fish.)

The 58th utterance was a text-only message, and the 71st message consisted of text and voice messages. The voice message of the 71st utterance mentioned “recommend eating fish.” In addition, user B sent a text message that summarizes the voice message. With this text message, users can surmise the contents of the voice message without listening to it.

In several cases, text messages, not voice messages, were used to express feelings. For example,

V [152] User B: (Voice message: uh...I don’t like it)

V [157] User A: >>[152] (Voice message: I can imagine your simpering face although you said “I don’t like it.”)

[162] User B: (´・ω・`)

The 162nd message is a Japanese-style face mark (emoticon) that expresses a downhearted feeling. It is generally said that voice is superior to text in expressing feelings. Indeed, this should be true in, at least, face-to-face communications. However, this generalization might not be true in the situations encountered by our subjects. It would actually be difficult to express “only” feelings without saying anything in non-face-to-face situations; text can nicely supplement voice to express only feelings.

#### 4. Conclusions

We attempted to extend multithreaded communications from single-modal to multimodal means. For this purpose, we developed a multithread-oriented multimodal chat system named “ChaTELex.” ChaTELex has particular functions for specifying a related message and receiver(s) to maintain multiple topic threads. In addition, ChaTELex allows a user to send an utterance that

includes text only, voice only or both text and voice.

We conducted pilot experiments using ChaTELex to explore possibilities and features of multithreaded multimodal communications. Based on the results of experiments, we confirmed it is possible to converse in multiple topic threads using multiple modalities, i.e., text and voice. Link information that shows semantic relations among messages plays a very important role in maintaining multiple topic threads. In addition, we found that ChaTELex could maintain many concurrent topic threads for a long time compared to a more limited baseline system. This was attributed to ChaTELex's functions for listening to the related utterances.

The results suggested two advantages of using text and voice. First, text can be used naturally to show a summary of a voice message. Furthermore, when a text-included message is related to a voice-only message, the text indicates the contents of the related voice message. Thus, without using speech recognition systems, we would be able to retrieve the voice messages; this improves the reusability of voice communication logs. As a second advantage, by using text emoticons, it becomes possible to convey intricate feelings that are difficult to express in a voice-only message. Accordingly, we can communicate feelings in a much richer way by using multiple modalities.

### Acknowledgements

This research was partially supported by the Grant Program of the Research Foundation for the Electrotechnology of Chubu, 2007, and by a JAIST research grant for in-house research projects, 2007.

### References

- Rhyll, V.: 2001, Applying membership categorization analysis to chat-room talk, *How to Analyze Talk in Institutional Settings: A Casebook of Methods*, McHoul, A. and Rapley, M. (eds.), Continuum, London, pp. 86-99.
- Winiacki, D. J.: 2003, Instructional discussion in online education, Practical and research-oriented perspectives, *Handbook of Distance Education*, Moore, M. and Anderson, R.M. (eds.), Lawrence Erlbaum Assoc..
- Ogura, K., Ishizaki, M. and Nishimoto, K.: 2004, A Method of Extracting Topic Threads towards Facilitating Knowledge Creation in Chat Conversations, *Proc. of KES2004*, Part I, pp. 330-336.
- Ogura, K., Nishimoto, K. and Sugiyama, K.: 2007, A Novel Voice Communication Systems based on Analyzing Text-Based Chat Conversations for Multiple Topic Thread, *Proc. of the Second IASTED International Conference on HUMAN-COMPUTER INTERACTION*, pp. 49-54.