Automatic Synchronization between Local and Remote Video Persons in Dining Improves Conversation

著者	Noguchi Yasuhito, Inoue Tomoo
journal or publication title	Journal of Universal Computer Science
volume	22
number	10
page range	1418-1435
year	2016-10
権利	(C) J.UCS
URL	http://hdl.handle.net/2241/00145076

Journal of Universal Computer Science, vol. 22, no. 10 (2016), 1418-1435
submitted: 29/2/16, accepted: 30/9/16, appeared: 28/10/16 © J.UCS

Automatic Synchronization between Local and Remote Video Persons in Dining Improves Conversation

Yasuhito Noguchi
(University of Tsukuba, Tsukuba, Japan
noguchi@slis.tsukuba.ac.jp)

Tomoo Inoue
(University of Tsukuba, Tsukuba, Japan
inoue@slis.tsukuba.ac.jp)

Abstract: Asynchronous exchange of video messaging is a way to achieve time-shifted communication for the people who have difficulties to enjoy daily family communication in real time, because of time-zone differences or life-rhythm differences. However face-to-face communication and video messaging communication is significantly different. Since mealtime is the most common opportunity for daily family communication, it has been proposed to synchronize the video message with the viewer by changing its playback speed in dining situations to improve video messaging communication. This paper studies the influence of the synchronization method by means of Wizard of Oz (WoZ), and by means of the implemented prototype system. In the synchronization method, the dining progress of the video person is matched with that of the viewer by real-time meal weight detection. The lab study found that the synchronization via WoZ increased speech frequency, decreased the duration of switching pauses, and led to a higher ratio of eating actions immediately after verbal responses of the user. This indicated that a more active commitment of the user was observed. The prototype system with finer control of the video than WoZ also achieved comparable result in terms of questionnaire scores, indicating the feasibility of a videoconferencing system with such a function.

Keywords: Video-mediated Communication, Time-shifted co-dining, Synchronization, Table talk
Categories: J.3, J.4

1 Introduction

Remote video-mediated communication has been one of the major topics of CSCW to date. Video-mediated communication in daily life has become widely accepted recently after decades of office use by professionals. New demands have been raised. Remote video-mediated co-dining is one of them. People desire not only simply video chatting, but also eating with their family or close friends. With a video chat tool such as Skype and FaceTime, people can connect easily with each other through video. They can enjoy eating together in front of the always-on video chat tool. However, such video chat tools are not versatile. All of them only support real-time communication. To the close friends and families who are geographically far apart and/or who have time-zone differences, those tools cannot provide a communication opportunity. Instead, conventional video messaging is the best option for this time-shifted (asynchronous) environment.

To improve this situation, a video-mediated, time-shifted, co-dining method called KIZUNA [Otsuka et al. 12] has been proposed. In this method, the user can eat while viewing a video message which was recorded previously. Thus he/she can eat with his/her partner artificially. In face-to-face co-dining, people adapt their eating speed unconsciously to a co-dining partner's speed [Hermans et al. 12]. In video-mediated, time-shifted co-dining, the viewers do not feel as though they are eating together because the video person's eating speed has not changed. In this method, dining synchronization is achieved through the video speed being adapted to the viewer's eating speed. Several effects of this method have already been reported.

The method was confirmed to have made an improvement regarding the subjective impression of the video person's speech timing and the feeling of co-dining [Nawahdah and Inoue 13]. Also, this method increased the viewer's utterance and turn-taking frequency [Inoue and Nawahdah 14]. The method is rather simple but seems to produce interesting effects on the user. Yet the analysis of the user's behavioral change was in only its initial stage from a few simple measures. Why the user's subjective impression was improved is not yet known. Thus, in this paper, further investigation into what is occurring with the user is done.

In co-dining, with the KIZUNA method, the user's remaining food is checked, and video speed is adapted. To check accurately, the experimenter has checked the user's food remains by hand (WoZ method). However, there is no person to check the food remains in an actual dining scene. Thus, implementation of the system for time-shifted co-dining has been put in place, as has comparing the co-dining with the system to the co-dining with the WoZ method. Whether or not to synchronize the dining progress has also been investigated.

2 Related Work

2.1 Conversation During a Meal

A few studies have analyzed conversation during a meal. For example, equalization tendency of utterance and gesture was observed when multiple participants talking over meal were compared with participation without a meal. This study suggested co-dining was useful when various opinions are called for from diverse people, especially from non-talkative people [Inoue and Otake 11]. Another study investigated the table talk with multiple participants focusing on their roles as a speaker and a hearer. This study suggested the participants decided when to eat depending on the degree of engagement with the conversation, resulting in cooperative and fluent table talk. In the study, it was observed that the speaker tended to eat soon after his/her utterance whereas the hearer tended to eat soon after his/her response, and this way the table talk was coordinated fluently [Tokunaga et al. 14]. These studies analyzed communication behavior in face-to-face conversation over a meal.

Situations where table talk occurs remotely with mediated technology are documented, but fewer studies are found, perhaps, because this situation is only recently emerging. One such study investigated the difference of face-to-face co-dining and remote co-dining. In the study, the visibility of a meal in remote co-dining was effective, as per the suggestion, to make it more closely resemble face-to-face dining [Furukawa and Inoue 13].

All the studies address conversation over a meal at the same time. No study, except this proposal, as described in the introduction, addresses conversation during a meal occurring at different times.

2.2 Co-dining Support System

The development of co-dining support systems seems to have been rather advanced compared to the analysis of co-dining or table talk. Accenture introduced a tele-dining prototype called the Virtual Family Dinner that would allow a remote family to dine together through a video connection. The prototype was essentially a videoconferencing system which was highly automated and easy to operate by targeting people with limited knowledge of technology, such as the elderly. The system monitors the site and when it detects a meal dish on the table, it goes through a list of contacts, trying to reach one who is available for a chat [Accenture 06]. The advertising agency Wieden+Kennedy's Amsterdam office produced a website, Virtual Holiday Dinner, enabling scattered friends and family to have a dinner party of up to five people via Skype. Guests can call into the dinner, and their faces are shown on the displays placed at the heads of models physically sitting around a dining table. The models are equipped with video cameras, so each guest can look around the dining table from the respective model's viewpoint by moving his/her head [Wieden+Kennedy 10].

A system called CoDine consists of a dining table embedded with interactive subsystems that augment and transmit the experience of communal family dining. CoDine connects people in different locations through shared dining activities, such as gesture-based screen interaction, mutual food serving, ambient pictures on an animated tablecloth, and the transmission of edible messages [Wei et al. 11].

These systems only achieve co-dining experiences at the same time.

2.3 Video-Mediated Time-Shifted Communication Support System

There are systems for supporting video-mediated time-shifted communication. They essentially exchange video messages in different settings. Asynchronous video messages were employed to support interpersonal relationships in separated families. The recipient viewed the video messages asynchronously, creating a non-stressful, continuous line of communication. This was believed to enhance the connectedness and intimacy between separated family members [Zuckerman and Maes 05]. Tang et al. introduced a system enabling a distant person to contribute to a workplace meeting by pre-recording comments to be played during the meeting when needed. The conducted field experiment showed that most of the recorded messages were played in the meetings, while a lesser percentage of the messages generated in the meeting were reviewed by the distant person [Tang et al. 12].

All these systems, however, only used the video without processing. The possibility of media processing for enhancing communication has not been explored. In contrast, this proposal explores media processing for the purpose of enhancing communication.

3 KIZUNA

3.1 Concept

While a person is having a meal, he/she is videotaped. The video will be sent to the partner. When the other person has a meal at a different time, the video is played back in front of him/her. This way, the basic environment of video-mediated time-shifted pseudo communication, which is a simple exchange of video messages, is established. In this proposed method, the playback speed of the video is adaptively controlled so that the dining progress of the video person is about the same as the one of the viewer in front of the video (Figure 1).

In face-to-face co-dining, dining synchronization is an important factor. In most cases, people start eating at the same time that their partner starts eating. It is known that eating behavior is influenced by the co-dining partner's behavior, and that people have a larger meal when eating alone [Castro and Brewer 92] [Patel and Schlundt 01]. Food amounts synchronize between co-diners [Conger et al. 80] [Herman et al. 03] because people adapt their eating speed unconsciously to a partner's speed [Hermans et al. 12]. Thus, in face-to-face co-dining, dining synchronization is occurring. In video-mediated, time-shifted, co-dining, dining synchronizing does not occur because the video person's eating speed does not change. In this method, dining synchronization is achieved by the video speed being adapted to the viewer's eating speed. Dining synchronization is performed by an approach in which a difference of the dining progress (DDP) between the viewer and the video person is kept within a threshold amount.

Several effects of this method have been reported already [Nawahdah and Inoue 13] [Inoue and Nawahdah 14]. The analysis of the user's behavioral change was studied in its initial stage only from a few simple measures. However, why the user's subjective impression was improved is not known. Thus, in this paper, further investigation is applied regarding the motives of the user. To check the co-diner's food remains accurately, the experimenter has checked the user's food remains in the experiment. However, there is no person to check the food remains in an actual dining scene. Thus, this investigation implemented the system for time-shifted co-dining, and compared co-dining with the system to co-dine with the WoZ method. Researchers for this project have investigated the system about whether or not to synchronize the dining progress.

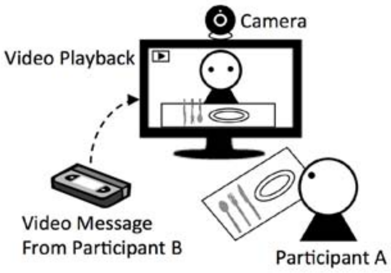

Figure 1: Conceptual Diagram of Time-Shifted Co-Dining

3.2 Co-Dining Synchronization with the WoZ Method

The assumed procedure is described as follows. A meal plate is put on a weight sensor. An USB camera is set next to the weight sensor. The image of the weight sensor's scale is sent to the other place. Another person can read a value of the scale with the image of the camera. He/she also can change the video speed every 1 minute. When more than 5% difference is found between the food remains from the video and from the user, the playback speed is changed to decrease the difference. The video playback speed is set to 0.7 times for slow playback and 1.5 times for fast playback, compared to 1 for standard playback. These were determined by the impression survey in advance so that the video could be watched without any serious problems.

3.3 Co-Dining Synchronization with the System

How to synchronize with the system is described in this section. A software application for the experiment, as described below, has been implemented. The software was written in Microsoft Visual C++. The software consisted of a module for checking DDP, a module for playing a video message, a module for videotaping a user, and a module for controlling the system. Each module runs in parallel while a user participates in co-dining. The system's work-flow is shown in Figure 2. A setting file and metadata is read ahead of co-dining. While co-dining, DDP is checked, the message video is played, and the viewer is recorded. After the meal, a period of the video person's speaking is recognized, and the data is saved as metadata. Finally, the recorded video with the metadata is sent to the co-dining partner.

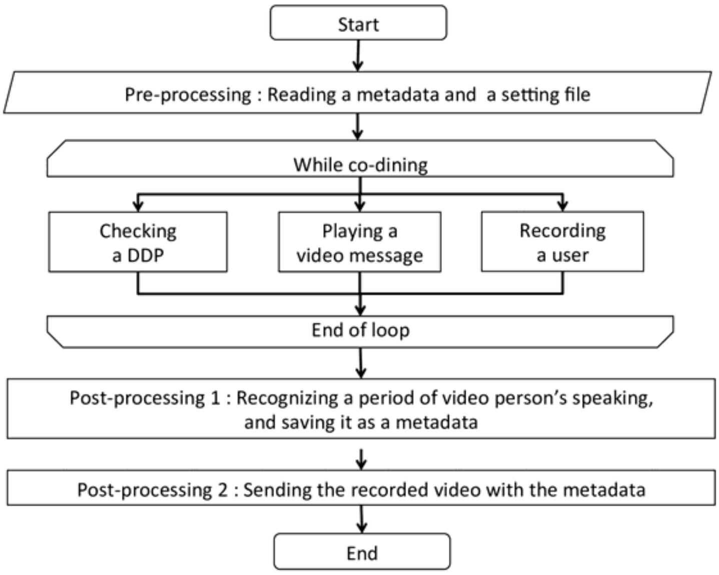
Start
Pre-processing : Reading a metadata and a setting file
While co-dining
Checking a DDP
Playing a video message
Recording a user
End of loop
Post-processing 1 : Recognizing a period of video person's speaking, and saving it as a metadata
Post-processing 2 : Sending the recorded video with the metadata
End

Figure 2: System's Work-Flow

3.3.1 Module for Checking DDP

This module has a function which checks the DDP. The meal plate is put on a weight sensor. An USB camera is set next to the weight sensor. The image of the weight sensor's scale is sent to the system. To measure the weight of the meal plate, this module carries image processing to the video of the scale. The remaining food is determined by subtracting the weight of the meal plate from the plate weight which was measured and input beforehand.

The results have been analyzed for the meal remains of 4 video of co-dining experiences. The average video length was 532 seconds. Two experimenters each labeled the meal remains when the user ate the food. Reproducibility of the two experimenters' labels is revealed (k=0.653 (n=153)) by kappa statistics [Kundel and Polansky 03]. The average RMS of the gap between the result of the labeling and the log data from the system was 7.8g (s.d. = 7.5g). The percentage of average error for the meal weight 300-500g was 1.5-2.5%. The precision is within the allowable range for this system. The analog device was used for checking the weight of the food. Both the analog sensor and digital sensor are acceptable for the purpose method if the device can measure the weight.

3.3.2 Module for Playing a Video Message

This module reads and plays the video message as a function. Another function is that it changes the speed of video playback. The libVLC API of VLC Media Player provided by VideoLan was used for playback and video speed controlling.

3.3.3 Module for Recording a User

This module records the user with an USB camera. This module also outputs the video written in the WMV format. The DirectShow provided by Microsoft was used to record the video.

3.3.4 Module for Controlling System

This module controls the whole system as one of its functions. The system is started when the “S” key on the computer is pressed by the user. This module receives the value of the initial meal weight from the module for checking the DDP when the system starts. At the same time, this module sends an order to play the video message. This module also receives the video person's dining progress data as metadata. At dinner time, this module receives DDP every second. The user's food remains rate is the percentage of the value passed from the module for checking the DDP in the initial meal weight. The video person's food remains rate is input from the metadata. This module gives an order to change the playback speed when the DDP is more than 5%. The video playback speed was set to 0.7 times for slow playback and 1.5 times for fast playback compared to 1 for standard playback. The playback speed was changed to decrease the DDP.

3.3.5 Software for Specifying the Speech Period of the Video Person

The system can control the video speed every 1 second. However, when the pitch of the video person's voice is changed because the playback speed of the video was changed, the user may sense something strange. The intelligibility of speech that controlled the speed was studied [Garvey 53]. When the rate of the standard speed is set to 100%, the variation of 30% influences the intelligibility a little. On the other hand, it's reported that the intelligibility becomes bad in the case of more than 70% [Daniloff et al. 68]. The video speed is controlled between x0.7 and x1.5, not to reduce the intelligibility of the viewer. However, the viewer might feel that the voice of the other person is unnatural to a certain degree, in case of controlling the speech speed, like this experiment [Uchida 05]. Therefore, the system controls the speed of the video, expecting the time when the video person speaks.

The software, which specifies the speech period, is carried out after co-dining (Figure 2). The system judges the video person's speech period by the sound pressure level of the video. First, the video file is converted to the sound file. The system calculates the RMS (Root Mean Square) value of the sound file every 1 second. If the RMS value is larger than the threshold value, the system judges that the video person is speaking. The sound file is written in a WAV format (PCM, 16bit, sampling rate: 441,00). The WAV file's sound data are signed integer type (from -32768 to +32767). The system sets a peak of the absolute value of the WAV file data (32768) as the standard value, and calculates the sound pressure level of the video. The result of the sound pressure level every 1 second was compared with the threshold value. If the result of the sound pressure level is larger than the threshold value, the system judges that video person is speaking at the 1-second period.

The results of specifying the speech period of 4 video messages have been analyzed. The average of the video length was 548 seconds. The video was labeled in terms of speech with a video annotation tool. The reproducibility of two experimenters' labels was revealed (k=0.823 (n=52)) by kappa statistic. On the other hand, the system calculated the sound pressure level from a WAV file of the video message. If the sound pressure level was larger than the threshold value (-88.1dB), the 1-second is judged as a speech period. The concordance rate between the result of labeling and the output of the system was 98.5%. The results of labeling and the output of the system almost matched.

4 Experiment

Two comparisons were conducted in this experiment. There are three conditions: the WoZ-sync condition wherein the pseudo co-dining is used in adaptive video playback speed with the WoZ method (WoZ-sync condition); the non-sync condition wherein the pseudo co-dining is used in normal video playback speed; and finally the system-sync condition wherein the pseudo co-dining is used in adaptive video playback speed with the system. To investigate the effect of synchronizing the dining progress of the video person in time-shifted co-dining, the WoZ-sync condition was compared with non-sync condition (Comparison 1). To investigate the effect of the method for dining synchronization, the system-sync condition was compared with the WoZ-sync condition (Comparison 2).

4.1 Video Message for the Experiment

In the experiment, the video message from the dining partner should be the same, and thus it was recorded while an actor (experiment cooperator) ate and talked according to the predetermined scenario shown in Table 1. Because the assumed co-dining situation, which is the most common situation, is with close people, such as family members, friends, and colleagues, the participant of the experiment should be a friend of the actor. As this constraints the number of participants, 5 video messages by 5 different actors were prepared. All 5 videos used the same scenario, but used different languages (2 Japanese, 2 Chinese, and 1 Arabic). Three video messages were used in the WoZ-sync condition and in the non-sync condition. The other 2 video messages were used in the system-sync condition. Each video included a single person that was watched by a single participant in this experiment. The actors were instructed to have a meal just like in daily life. The sentences of the scenario were from questions and answers, the corresponding time was set considering the time to respond, and a 400g plate of curry with rice and a soft drink was used as the meal in the videos. The video lengths resulted in about 9 minutes. In the WoZ-sync condition and in the non-sync condition, the meal progress in each video was measured and recorded every minute before the experiment. In the system-sync condition, the meal progress in the video was measured and recorded every second with the system before the experiment.

(mm:ss)	Questions and Comments
Start	Q) Hello, how are you today?
00:45	The weather here is so nice today, I like the summer season.
01:30	Q) Do you like your meal?
02:15	Delicious, I like curry rice.
03:00	Q) By the way, what’s your favorite food?
03:45	Personally, I like the (Italian) food a lot.
04:30	Q) Where do you live?
05:15	I like (Tsukuba) city. It’s safe, clean and the people are so friendly.
06:00	Q) Do you have any plan for the summer vacation?
06:45	I like the sea a lot, so most probably I will go to a beach and have some relaxed time.
07:30	Q) Which country would you like to visit?
08:15	Nice, I like to visit (Italy). I want to go there to eat (Italian) food.
End	Thank you. I am looking forward to meeting you again in the next video.

Table 1: Scenario of the Recorded Video Message

4.2 Setup

The experimental session was conducted in the booth of the lab so that the environment was controlled across the conditions. The actual scene is shown in Figure 3. The experimental system consists of a PC, a display, a speaker, two USB cameras, and an analog weight sensor (0-500g) for all conditions. USB camera 1 was used to record the viewer's facial expressions, gestures, and responses, which could become a response video message. USB camera 2 was used to read the weight of the meal, which the experimenter used for controlling the video. In the WoZ-sync condition or the non-sync condition, the experimenter reads a value of the weight sensor with USB camera 2. In the system-sync condition, the system reads a value of the weight sensor by image processing of the image from USB camera 2.

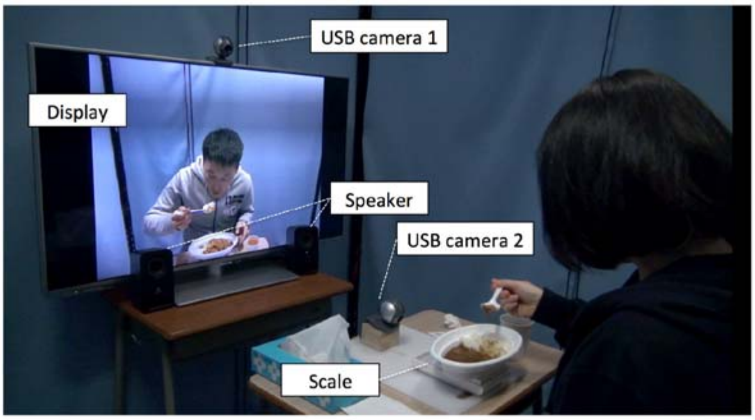

Figure 3: Actual Scene of the Experiment

4.3 Participants

There were 36 university students - 15 male and 21 female - participating in the experiment. They were divided in 3 groups. None of the participants had experienced this type of experiment before.

4.4 Meal

A plate of curry with rice and a soft drink was used as the meal in each scenario. Each participant chose the size of the meal, from 300g or 500g, which was different from 400g in the videos. This was intended to represent the meal difference in the real world within the limitation of a controlled experiment. The weights were determined after investigating similar commercial products. Compared with 400g, 300g represents a 25% decrease and 500g represents a 25% increase. The meal consisting of 300g was chosen by 21 participants, and the 500g meal was chosen by 15 participants.

4.5 Procedure

After the briefing and the information consent for the experiment, the participant was guided to the booth and was instructed regarding the scenario of the setup. He/she was instructed that the person in the video would talk, that his/her dining and talking was recorded, and would be watched by the video person later. He/she was asked to behave just like in his/her daily life. He/she was not informed of the experimental conditions. The experimenter then left the booth and started the video playback which was the indication of the beginning of the session. In the WoZ-sync condition, the video speed was adapted with the WoZ method. In the system-sync condition, the video speed was adapted with the system. In the non-sync condition, the video speed was not changed. After the session, he/she was asked to fill out a questionnaire.

5 Results

5.1 Comparison Between the WoZ-sync Condition and Non-sync Condition (Comparison 1)

The recorded video was analyzed in terms of the participant's communication and eating behavior. The analyzed part of the video included material from the beginning of the video message to the end of the video message, or from the beginning of the video until the time the participant finished a meal. The average length of the video was 8.8 minutes (s.d. = 1.2). Regarding playback time, 28% of the total playback time was not in the standard speed, where 15% was in slow playback, and 13% was in fast playback.

5.1.1 Descriptive Values

The video was labeled in terms of speech and eating using a video annotation tool. The measures were:

- Speech frequency: The number of words in a minute
- Speech length: The average length of speech
- Switching pause of the participant: The time needed for the participant respond to the video person's speech
- Switching pause of the video person: The time until the video person speaks after the participant's speech
- Overlapping frequency: The number of overlapping words between the video person and the participant in a minute
- Response rate: The rate of the participant's response to the video person's speech
- Eating frequency: The amount of eating within a minute.

Here, any utterance longer than 1.5 seconds was identified as speech, and the action of taking the food and bringing it to the mouth was identified as eating. The result of each measure for both conditions and the statistical significance of those differences is shown in Table 2. This information was compared with the Mann-Whitney U test because the data of speech length, overlapping frequency, and response rate was not normally distributed. Similarly, the Mann-Whitney U test was used because the variances of speech frequency (F(1, 22) = 4.09, p<0.05) and switching pause of the participant (F(1, 22) = 5.49, p<0.05) was unequal. An independent t-test was used for comparison of the other data. The speech frequency tends to be larger in the WoZ-sync condition than in the non-sync condition (U=42, Z=-1.732, p=0.083). The switching pause of the participant was significantly shorter in the WoZ-sync condition than in the non-sync condition (U=22.5, Z=-2.859, p=0.004).

Measures	WoZ-sync.	Non-Sync.	p-Value
Speech frequency (times/minute)	3.45	2.43	*0.083
Speech length (sec.)	3.49	3.11	0.908
Switching pause (Participant) (sec.)	1.57	2.89	***0.004
Switching pause (Video person) (sec.)	16.9	20.6	0.313
Speech overlap (times/minute)	0.40	0.27	0.885
Response rate (%)	85.2	78.7	0.487
Eating frequency (times/minute)	3.07	2.81	0.446

***: $p < 0.01$, *: $p < 0.10$

Table 2: The Average Data per Condition Result

5.1.2 Relation Between Speech and Eating

The time-shifted communication in this setup is pseudo communication where the participant responds to the message of the video person. Because of the style that the participant watches the video, it was common that the participant became a hearer of the video person.

In a study that addressed the behavior of a hearer in face-to-face triad table talk, it was reported that the hearer adjusted the eating timing according to the degree of engagement in the conversation, which contributed to being cooperative with table talk. The degree of the hearer’s engagement to the conversation was defined as high when the hearer was directly addressed by the speaker.

In this situation, it was observed that the hearer responded to the speaker before eating. In other words, the conversation gained priority over eating. This eating action, adjacent to the response, was regarded as distinctive action of attentive listening of the hearer [Tokunaga et al. 14].

The degree of engagement could be measured by investigating the eating action adjacent to the response, according to this study (Figure 4). As our experimental setting was triad table talk, the participant was always the direct addressee of the speech. The assumption, however, was that there could be a different degree of engagement to the conversation in different triad conversations.

The study compared this “adjacent eating” between the conditions. All the first eating actions the participant took during or after the video person’s speech were investigated, whether it was adjacent eating or not. The adjacent eating is the eating action after the response within certain limited time. As for the response, verbal responses, such as answering the video person’s question and back-channeling, are included. Nonverbal responses, such as turning the head, gazing, and nodding, are not included because this measure intends to determine in which case speech takes priority over eating.

Whether the response was corresponding to the video person’s speech, and thus was acceptable as a response or not, was determined by the two experimenters, independently. As for the certain limited time within which eating action should follow after the response, adequate maximum time is regarded as twice the standard deviation of the stroke action of eating, which can be acceptable as the time-lag between intention and behavior of eating [Tokunaga et al. 14].

In this experiment, the stroke action was from scooping the food by a spoon to carrying it to the mouth. The total number of eating actions was 706. The average stroke length was 1.6 seconds, with the standard deviation of 0.5 seconds. Thus, the limited time in this experiment was 2.6 seconds (=1.6+2*0.5). Therefore, the precedence of the response within 2.6 seconds was examined for all of the above sampled eating.

Table 3 shows the result. In the WoZ-sync condition, the number of all the sampled eating was 142, where the number of adjacent eating was 114. The rate of adjacent eating was 83%. In the non-sync condition, the number of all the sampled eating was 126, where the number of adjacent eating was 87. The rate of adjacent eating was 66%. These examples were compared with the Mann-Whitney U test because the variance of the data was unequal (F(1, 22) = 4.07, p<0.05). Significant difference was found in the rate of adjacent eating between the conditions (U=35.5, Z=-2.115, p=0.034). The adjacent eating was taking place more often in the WoZ-sync condition. This suggested a higher degree of engagement to conversation in the WoZ-sync condition.

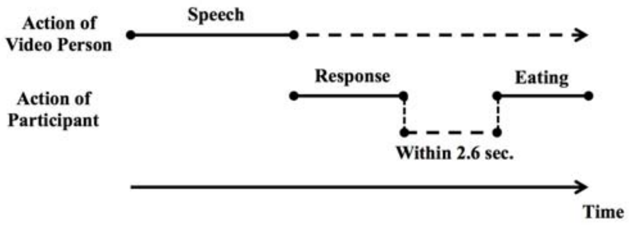

Figure 4: Adjacent Eating

	WoZ-sync.		Non-sync.		p-Value
	Adjacent	Non-Adjacent	Adjacent	Non-Adjacent	
Number of Samples	114	28	87	39	**0.034
Rate	82.6%	17.4%	66.0%	34.0%	

**: p < 0.05

Table 3: Adjacent Eating Rate

5.2 Comparison Between the System-sync Condition and the WoZ-sync Condition (Comparison 2)

The comparison between the system-sync condition and the WoZ-sync condition used a questionnaire. The DDP was also evaluated in both conditions. The average length of the video was 7.1 minutes (s.d. = 1.7). Regarding playback time, 41% of the total playback time was not in the standard speed, where 18% was in slow playback, and 23% was in fast playback.

5.2.1 Questionnaire

In the questionnaire, participants were asked to evaluate a series of statements according to the feelings they experienced during the time-shifted dining session. The questionnaire was written in English because all participants were able to answer using that language. The results of questionnaire are shown in Table 4. The Mann-Whitney U test was used for the comparison between both conditions. The principal aim of the study was to assess participants' communication with their partners, and to determine whether synchronizing the dining sessions would affect communication or not, using a selection of straightforward questions from related questionnaires [Inoue et al. 97] [Sellen 92] (C1-C6). The sense of the partner's presence while tele-dining is another important aspect to consider in assessing the proposed system. Synchronizing the dining sessions is believed to affect this sense of the partner's presence. This study sought to determine whether the proposed system would enhance this sense or not, using common questions from related questionnaires [Kies et al. 97] [Nakanishi et al. 11] [Ichikawa et al. 95] (P1-P6). All of the above statements were rated on a 9-point Likert scale, where 1 = strongly disagree, 3 = disagree, 5 = neutral, 7 = agree, and 9 = strongly agree. Comparison between both conditions was done by the Mann-Whitney U test. As the result, there is no significant difference between both conditions.

No.	Questionnaire	System-sync.	WoZ-sync.	p-Value
C1	I wanted to talk to the partner.	6.9	7.0	0.930
C2	I enjoyed talking with the partner while eating.	6.5	6.8	0.705
C3	The partner's talking distracted me from my meal.	4.5	4.6	0.857
C4	The content of the conversation was natural.	5.4	5.8	0.554
C5	The timing of the partner's delivery was natural.	5.5	5.5	0.929
C6	I could communicate with the partner naturally.	5.8	5.5	0.952
P1	I felt as if the partner and I were eating together in the same room.	6.0	6.5	0.312
P2	I felt distant from the partner.	5.0	4.7	0.361
P3	The partner's facial expressions were easy to recognize.	6.4	6.6	0.952
P4	The partner's gaze direction was easy to recognize.	6.0	6.3	0.858
P5	I was able to make eye-contact with the partner.	5.0	5.0	0.977
P6	The partner's gestures were easy to recognize.	6.8	6.8	0.857

Table 4: The Result of the Questionnaire

5.2.2 Dining Progress

The actual dining progress with and without synchronization (system sync) is shown in Figure 5. The "Actor speech" area shows the period when controlling playback speed was prohibited due to actor's speech. The "Fast" area shows the period when the video was played fast. The "Slow" area shows the period when the video was played slowly. The "Participant" line reveals the food remains of a participant. The "Actor sync" line reveals the food remains of the video person in the synchronizing condition. The "Actor non sync" line shows the simulated food remains of the video person if it was in the controlled condition. The label on the figure reveals a difference between the value of the "participant" and the value of the "Actor sync".

Similarly, the actual dining progress with and without synchronization (WoZ sync) is shown in Figure 6. The dining progress of both the participants were each slower than the dining progress of the video persons, and the videos were played slowly. Fast/slow playback speed in the synchronizing condition helped decreasing DDP.

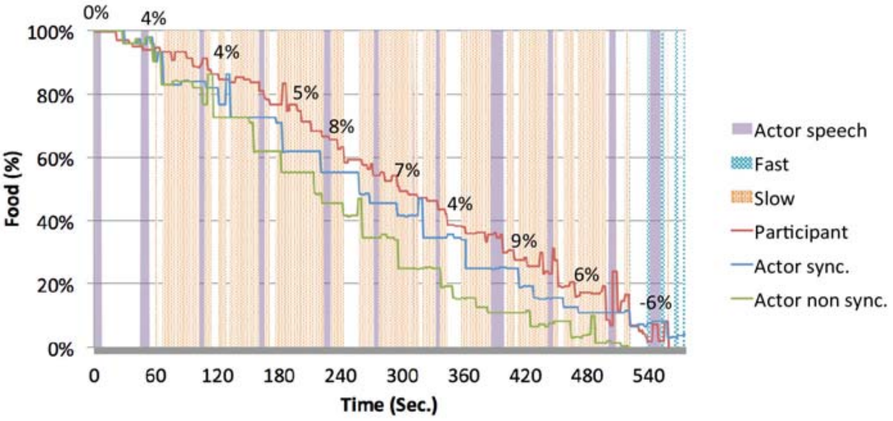

Figure 5: The Actual Dining Progress with and without Synchronization (system sync)

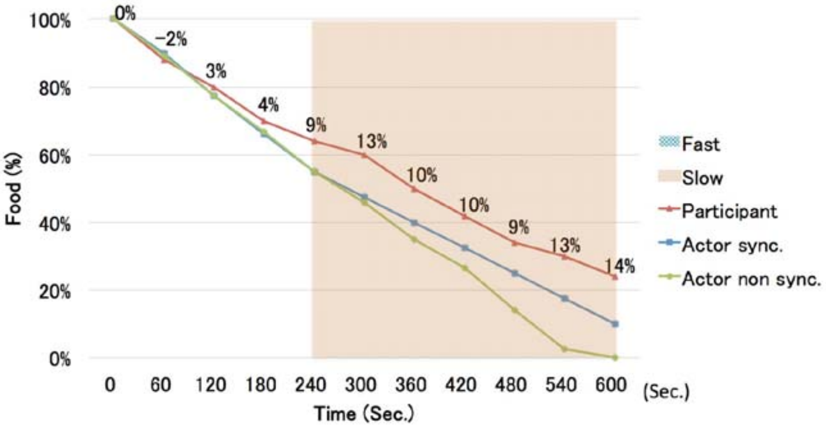

Figure 6: The Actual Dining Progress with and without Synchronization (WoZ sync)

6 Discussion

6.1 Behavioral Change in Synchronizing the Dining Progress

Synchronizing the dining progress had the effect of increasing speech frequency and decreasing the length of switching pause of the participant, as shown in Table 2. It indicates that the participant involves in conversation more actively when the dining progress is synchronized. Synchronizing the dining progress also induces more adjacent eating after the response to the video person, as shown in Table 3. This behavior indicates that the participant feels more engaged with the conversation with the video person.

6.2 Unaware Behavioral Change Causing Better Subjective Feelings

It was reported that synchronizing the dining progress increased the feeling of eating together in the same room, and the feeling of naturalness of the video person's speech timing [Nawahdah and Inoue 13], despite the fact that the video person's speech timing was not controlled. In the beginning, this subjective feeling was unknown. First, the hypothesis was that this might be related to the difference of speech overlaps between the video person and the participant. If there are frequent speech overlaps, it could be recognized as unnatural. The result, however, was negative to this hypothesis. The experiment found no significant difference regarding the speech overlap, as in Table 2.

The next hypothesis was that this might be related to the difference of switching pause of the video person. If the switching pause of the video person happened to be shorter in the WoZ-sync condition as a result of video playback speed control, it could be recognized as more natural because the switching pause in face-to-face conversation is shorter. The result was again negative to this hypothesis. No significant difference was found regarding the switching pause of the video person, as in Table 2.

In contrast with these initial hypotheses that were negated, a significant difference between the conditions was found in the switching pause of the participant, not the video person. The switching pause of the participant was significantly shorter in the WoZ-sync condition. This shorter switching pause could make the participant feel more natural regarding the speech timing. Also, more adjacent eating, which was also the behavior of the participant, was observed in the WoZ-sync condition. This could make the participant feel more engaged in the conversation. Surprisingly, these results revealed that the participant adjusted when to talk and when to eat cooperatively with the video person's recorded behavior. The participants were not aware that they changed their behavior, which caused their subjective feelings better.

6.3 On the Other Synchronization

The effect of synchronizing non-verbal information has been investigated. In this study, the system has used the information about user's meal remains in co-dining. However, other non-verbal information could be useful in a similar way. As an example, if one was to attempt to sleep using the video message for sleeping together with family, one could sleep well if the breath timing is synchronized. Behavior synchronizing could influence time-shifted communication like the proposed method.

6.4 The Implemented System

The system measured food remains with high accuracy. The system could also recognize the period of video person's speaking with high accuracy. In this experiment, the meal plate was set on a weight measure scale. However, the method cannot work well when the meal is something eaten by hand, such as a sandwich. In that case, image processing through a camera should be used for measuring food remains. The video messages used in this experiment were recorded in a quiet environment, where there was no noise. The system could not specify the speech period if the video was recorded in a noisy place. Improvement is needed for adaptation to various scenes. The effect of co-dining synchronizing could be strong if accuracy of the software increases.

7 Conclusion

In this paper, the effect of dining synchronizing for video-mediated, time-shifted communication has been explored experimentally. In this result, the proposal of synchronizing the video with its viewer may be one way to improve this type of communication. The experiment proved that the WoZ synchronization increased speech frequency and decreased the duration of switching pauses of the user. Also observed was a higher ratio of eating actions immediately after verbal responses, which indicated a more active commitment of the user.

Furthermore, the experiment implemented the time-shifted co-dining system, which checks the food remains of a user every 1 second. The evaluation of the system revealed that the system worked almost accurately. From the result of comparison between the dining synchronizing with the system and the dining synchronizing with the WoZ, the implemented system can synchronize the user's dining speed to the speed of video person's dining in time-shifted co-dining, as in the case of using the WoZ method. The implemented system is comparable to the method by hand. This paper highlights one design of the asynchronous co-dining system, which indicated more active commitment of the user.

Acknowledgements

This study was financially supported by JSPS Grant-in-Aid for Scientific Research 26330218 and 15K00888.

References

[Accenture 06] Accenture: "VIRTUAL FAMILY DINNER"; (2006), http://gizmodo.com/accenture-virtual-family-dinner/

[Beasley et al. 72] Beasley, D. S., Schwimmer, S., Rintelmann, W. F.: "Intelligibility of Time-Compressed CNC Monosyllables"; J Speech Hear Res., 15 (1972), 240-350.

[Castro and Brewer 92] Castro, J. and Brewer, M.: "The amount eaten in meals by humans is a power function of the number of people present"; Physiology and Behavior, 51, 1 (1992), 121-125.

[Conger et al. 80] Conger, J., Conger, A., Costanzo, P., Wright, K., and Matter, J.: "The effect of social cues on the eating behavior of obese and normal subjects"; J Pers, 48, 2 (1980), 258-271.

[Daniloff et al. 68] Daniloff, R. G., Shriner, T. H., Zemlin, W. R.: "Intelligibility of Vowels Altered in Duration and Frequency"; J. Acoust. Soc. Am., 44 (1968), 700-707.

[Furukawa and Inoue (2013)] Furukawa, D., Inoue T.: "Showing meal in video-mediated table talk makes conversation close to face-to-face"; IPSJ (Information Processing Society of Japan) Journal, 54, 1 (2013), 266-274.

[Garvey 53] Garvey, W. D.: "The Intelligibility of Speeded Speech"; J Exp Psychol., 45, 2 (1953), 102-108.

[Herman et al. 03] Herman, P., Roth, D., and Polivy, J.: "Effects of the Presence of Others on Food Intake: A Normative Interpretation"; Psychological Bulletin, 129 (2003), 873-886.

[Hermans et al. 12] Hermans, R., Lichtwarck-Aschoff, A., Bevelander, K., Herman, P., Larsen, J., and Engels, R.: "Mimicry of Food Intake: The Dynamic Interplay between Eating Companions"; PLoS ONE, 7, 2 (2012).

[Ichikawa et al. 95] Ichikawa, Y., Okada, K., Jeong, G., Tanaka, S., and Matsushita, Y.: "MAJIC Videoconferencing System: Experiments, Evaluation and Improvement"; Proc. ECSCW'95, Springer, Stockholm (1995), 279-292.

[Inoue et al. 97] Inoue, T., Okada, K., Matsushita, Y.: "Integration of face-to-face and video-mediated meetings: HERMES"; Proc. GROUP '97, ACM Press, Phoenix (1997), 405-414.

[Inoue and Otake 11] Inoue, T., Otake, M.: "Effect of meal in triadic table talk: Equalization of speech and gesture between participants"; Transactions of Human Interface Society, 13, 3 (2011), 19-29.

[Inoue and Nawahdah 14] Inoue, T., Nawahdah, M.: "Influence of dining-progress synchrony in time-shifted tele-dining"; Proc. CHI '14 Extended Abstracts on Human Factors in Computing Systems, ACM Press, Toronto (2014), 2089-2094.

[Kies et al. 97] Kies, J. K., Williges, R. C., and Rosson, M. B.: "Evaluating desktop video conferencing for distance learning"; Computers and Education, 28, 2 (1997), 79-91.

[Kundel and Polansky 03] Kundel, H. L., Polansky, M.: "Measurement of observer agreement"; Radiology, 228, 2 (2003), 303-308.

[Nakanishi et al. 11] Nakanishi, H., Kato, K., and Ishiguro, H.: "Zoom cameras and movable displays enhance social telepresence"; Proc. CHI'11, ACM Press, Vancouver (2011), 63-72.

[Nawahdah and Inoue 13] Nawahdah, M., Inoue, T.: "Virtually dining together in time-shifted environment: KIZUNA design"; Proc. CSCW'13, ACM Press, San Antonio (2013), 779-788.

[Otsuka et al. 12] Otsuka, Y., Nawahdah M., Inoue T.: "Development of KIZUNA system capable of time-shifted co-dining communication"; Technical report of The Institute of Electronics, Information and Communication Engineers, 112, 75 (2012), 85-90.

[Patel and Schlundt 01] Patel, K. and Schlundt, D.: "Impact of moods and social context on eating behavior"; Appetite, 36, 2 (2001), 111-118.

[Sellen 92] Sellen, A. J.: "Speech patterns in video-mediated conversations"; Proc. CHI'92, ACM Press, Monterey (1992), 49-59.

[Tang et al. 12] Tang, J., Marlow, J., Hoff, A., Roseway, A., Inkpen, K., Zhao, C., Cao, X.: "Time Travel Proxy: Using lightweight video recordings to create asynchronous, interactive meetings"; Proc. CHI'12, ACM Press, Austin (2012), 3111-3120.

[Tokunaga et al. 14] Tokunaga, H., Mukawa, N., Kimura, A.: "Structure of Cooperative Communication Behavior During Table Talk: When do Hearers Eat and When do They Respond"; Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, 26, 4 (2014), 793-801.

[Uchida 05] Uchida, T.: "Impression of Speaker's Personality and the Naturalistic Qualities of Speech: Speech Rate and Pause Duration"; Japanese Journal of Educational Psychology, 53, 1 (2005), 1-13.

[Wei et al. 11] Wei, J., Wang, X., Peiris, R. L., Choi, Y., Martinez, X. R., Tache, R., Koh, J. T. K. V., Halupka, V., Cheok, A. D.: "CoDine: an interactive multi-sensory system for remote dining"; Proc. UbiComp'11, ACM Press, Beijing (2011), 21-30.

[Wieden+Kennedy 10] Wieden+Kennedy Amsterdam office: "Virtual Holiday Dinner"; (2010), http://www.digitalbuzzblog.com/wieden-kennedy-virtual-holiday-dinner/

[Zuckerman and Maes 05] Zuckerman, O., Maes, P.: "CASY: Awareness System for Children in Distributed Families"; Proc. IDC'05, ACM Press, Boulder (2005).