Virtually dining together in time-shifted environment: KIZUNA design

Conference Paper · January 2013

DOI: 10.1145/2441776.2441863

Mamoun Nawahdah
Birzeit University

Virtually Dining Together in Time-Shifted Environment: KIZUNA Design

Mamoun Nawahdah
Graduate School of Library, Information and Media Studies
University of Tsukuba
nawahdah@slis.tsukuba.ac.jp

Tomoo Inoue
Faculty of Library, Information and Media Science
University of Tsukuba
inoue@slis.tsukuba.ac.jp

ABSTRACT
Dining with a remote person requires that both participants are available at the same time to eat together. Because of time-zone differences and other such contingent factors, this condition can often be hard to fulfill. One solution may lie in time-shifted communication. A person can enjoy a meal while watching an earlier recorded video of a remote person's dining. However, in a time-shifted environment, achieving dining synchronization is a challenge. In this research, we propose a time-shifted tele-dining system (KIZUNA) enabling people to enjoy a meal together in a virtual environment. The system adapts the displayed video's playback speed to the difference in dining progress between the local and remote person. This is likely to enhance communication and increase enjoyment while dining. A validation experiment revealed that the proposed KIZUNA adaptation method enhanced diners' communication behavior, and significantly enhanced the perceived presence of the remote person, in comparison with conventional time-shifted tele-dining. This result suggests a promising future for the KIZUNA system.

Author Keywords
Tele-dining; time-shifted communication; videoconferencing; adaptation; social networking.

ACM Classification Keywords
H.5.3 Group and Organization Interfaces: Asynchronous interaction

INTRODUCTION
During meals, families have a chance to catch up on what has happened in their lives, and to strengthen the bonds that hold the family together. The busy lifestyles of our time, however, often make such family dining very difficult; especially with older people living independently, younger people living alone, and working people traveling more and/or working from afar [21]. In these cases, family members must often eat alone, which is a possible source of loneliness and unhappiness [5][16], in contrast to the pleasure and longer duration typical of social dining [6].

A variety of inexpensive videoconferencing technologies offer a decent solution for distant family members to maintain a sense of social connectedness, which is defined as "positive emotional appraisal, characterized by a feeling of staying in touch within ongoing social relationships" [20]. But when family members are located far away from each other, this often means they are also living in different time-zones, and this difference is likely to present further difficulties for communication between distant family members. The main difficulty lies in the misalignment of daily schedules between the two parties [1], and this is clearly apparent at mealtimes.

In this paper, we propose the "KIZUNA" system to overcome the problem of eating alone. The system is principally based on time-shifted communication involving the transmission of recorded video messages (Fig. 1). To achieve fine synchronization between local dining and the displayed recorded video, KIZUNA adapts the displayed video's playback speed to the difference in dining progress (DDP) between the local and remote user. KIZUNA system usage will vary with circumstances; it can be used occasionally or frequently, by family members or friends. The system is currently under development, with numerous modules having already been completed and tested. The KIZUNA's adaptation algorithm was validated by an actual dining experiment using the Wizard of Oz (WoZ) system simulation technique. Utilizing WoZ, we replicated an envisaged future scenario for the KIZUNA system, and the experimental results revealed that the proposed KIZUNA adaptation method enhanced diners' communication behavior and significantly enhanced the perceived presence of the remote person, in comparison with conventional time-shifted tele-dining.

RELATED WORKS
Numerous studies have investigated ways to utilize information and communication technology in minimizing the gap between people living apart. In this section, we briefly review research related to people dining together despite being located in remote places and/or different time-zones.

Synchronous Dining
Family members and close friends tend to have meals together, but lack of time or proximity often makes this hard to

CSCW'13, February 23–27, 2013, San Antonio, Texas, USA.

Figure 1. User dines while watching a recorded video message.

achieve. To overcome this situation, the international consulting firm Accenture introduced a tele-dining prototype called the Virtual Family Dinner that would allow families to dine together in a virtual environment[1]. This prototype was essentially a videoconferencing system targeting people with limited knowledge of technology, such as the elderly. Thus, Accenture made it highly automated and easy to operate. 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 dinnertime chat.

The advertising agency Wieden+Kennedy's Amsterdam office produced a website called Virtual Holiday Dinner[2], 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 as the heads of models physically sitting around a dining table. The models are equipped with video cameras fitted with facial tracking software, so each guest can look around the dining table, 'from' the respective model, by moving his/her head.

The sense of coexistence among family members is likely to be affected when living apart. To maintain a heightened sense of coexistence among family members, a system called CoDine was introduced [25]. This system 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.

Dining together is also an enjoyable experience. In some cultures, meals are typically consumed in groups, usually made up of family members or close friends. Examples of popular group meals are the Chinese Hot-Pot, the Japanese Nabe, and the Arabic Mansaf. In this regard, remote group-meal communication was investigated to render such meals more enjoyable [2]. In the study (of the Chinese Hot-Pot meal), three factors essential to the group-meal experience were identified: interacting with food as a group, a central shared hot-pot, and a feeling that others are nearby.

All such previous research primarily focused on synchronous dining among a group of people living apart, assuming that all participants are available at the same time. Our research, in contrast, focuses on the cases where only one participant is currently available for dining.

Asynchronous Dining

When all participants are available at the same time, dining together can be achieved by a variety of videoconferencing systems. However when only one participant is available, things become more difficult. In this situation, asynchronous (time-shifting) collaboration techniques offer a general solution. Because users can receive the information they want when it is most convenient for them, such techniques have been widely employed in work settings, but they may lack a sense of immediacy. For example, in the distant-learning domain, Ocker et al. revealed that asynchronous collaboration is as effective as face-to-face collaboration in terms of learning, quality of solution, solution content, and satisfaction with the solution quality [18]. However, students were less satisfied with the asynchronous learning experience in terms of both the group interaction process and the quality of group discussion. This suggests that asynchronous interaction is satisfactory for some requirements, but not for others, such as affective satisfaction or social connectedness. Asynchronous video messages have also been employed to support interpersonal relationships in separated families [26]. Here, the recipient views the video messages asynchronously, creating a non-stressful, continuous line of communication. This is believed to enhance the connectedness and intimacy between separated family members. Inkpen et al. showed that despite enjoying face-to-face interaction, close friends used asynchronous video communication tools extensively to augment their existing relationships [9]. Time-shifting has also been considered for meetings in the workplace. Tang et al. introduced a system enabling a distant person to contribute to a meeting by pre-recording comments to be played during the meeting when needed [23]. The conducted field experiment showed that most of the recorded messages were played in the meetings; however, a lesser percentage of the messages generated in the meeting were reviewed by the distant person.

In the dining domain, Tsujita et al. proposed the system "CU-Later", to be used in time-shifted dining [24]. This system plays a recorded video of remote dining after a specific time-shift, when the local user is in front of a display placed on a dining table, enabling the local user to watch the video automatically when he/she is eating. As the video is played, the system records the local user's session as well, so that the remote user can watch the local user's session later on-rather like a video mail exchange with automatic playback and recording.

Social dining communication means more than just information exchange; it typically also involves affective satisfaction and social connectedness, and may thus be seen as a form of 'consummatory communication' (which involves the

[1]http://gizmodo.com/accenture-virtual-family-dinner/
[2]http://www.virtualholidaydinner.com/

sharing of experiences, emotions, knowledge, and/or opinions [3]). To support this type of communication, a sense of co-presence is often crucial. Thus, our proposed system addresses these more complex needs, and further explores the potential of video exchange, but is more focused on the social connectedness between remote users, based on the assumption that synchronizing the dining activities of remote participants may significantly enhance the users' dining experience.

KIZUNA SYSTEM

Figure 2 shows a conceptual illustration of the KIZUNA ('bond' in Japanese) system scenario. In this scenario two persons, local user A and remote user B, are dining together in a virtual environment, although they are physically dining in different spaces and/or at different times. First (Fig. 2 Session 1), a recorded video message from the remote user B is created while dining.

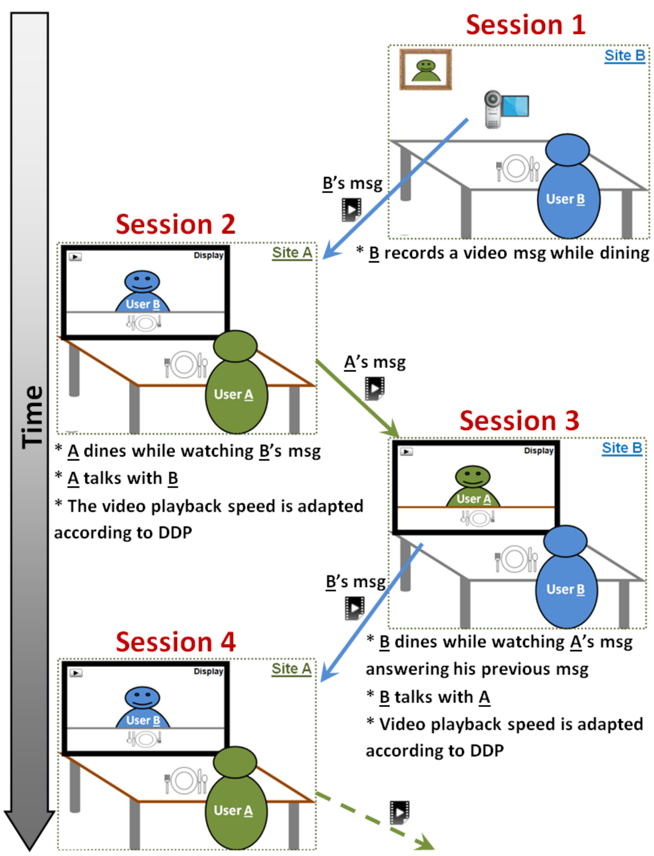

Figure 2. KIZUNA system scenario.

The system automatically starts working at site A when the local person A gets seated and some dishes are placed on the dining table (Fig. 2 Session 2). This increases the system-usage portability, especially for people with limited knowledge of the technology. The system starts playing the video of remote user B (recorder earlier) on a large display in front of the local dining table, at life-size scale. It also starts recording the local user A's dining and his/her reaction to user B's video message. Similarly, the system automatically starts working at site B when the local person B gets seated and some dishes are placed on the dining table (Fig. 2 Session 3). The system then starts playing the video of user A, and the same process is repeated. This automatic playback and recording of videos provides the illusion of co-dining for the users, though they may be dining at different times and/or in different places. We assume that merely watching the other user's video may not, by itself, arouse a sense of co-dining; but that the synchronization of both dining activities (not only the start and end time but also the entire process) may be significant. Specifically, this synchronization can be achieved by controlling the video playback speed according to the DDP between the local and remote user. In this way, the local user can enjoy the company of the remote user throughout his/her meal. Also in this way, we ensure that the local user can see the entire remote dining session while he/she is eating. These factors appear likely to enhance the dining experience.

System Design

Based on the system proposal, we defined the following set of design requirements to implement the KIZUNA system: **Dining detections**: dining status and progress are necessary information to control the system. This includes the following:

- User detection: the system must be highly automated and easy to operate. User presence at the dining table should activate the system and display the remote participant's recorded dining video. This will ensure high system usability and portability, especially for people with limited knowledge of the technology.
- Food detection: detecting a user at the dining table is not sufficient to ensure that a dining session is about to begin; the system must also detect whether food has been placed on the dining table or not.
- Amount of food detection: based on the above information, the system can assess the user's dining behavior, particularly the user's dining progress.

Video manipulation: in the proposed KIZUNA system, users can communicate through the exchange of recorded video messages. Therefore, the following video manipulations are required:

- Video recording: to record the local dining session.
- Video streaming and controlling: to control the displayed video's playback speed according to the DDP.

System Implementation

An experimental time-shifted tele-dining system is under construction based on the KIZUNA system proposal and relevant design requirements. The experimental site is equipped with the following: dining table, chair, large flat-panel display, two USB cameras, speakers, microphone, motion detection sensor, two spotlights, and computer. Figure 3 shows the preliminary system's physical workspace.

Figure 4 shows the system procedure flowchart, comprised of the following major modules:

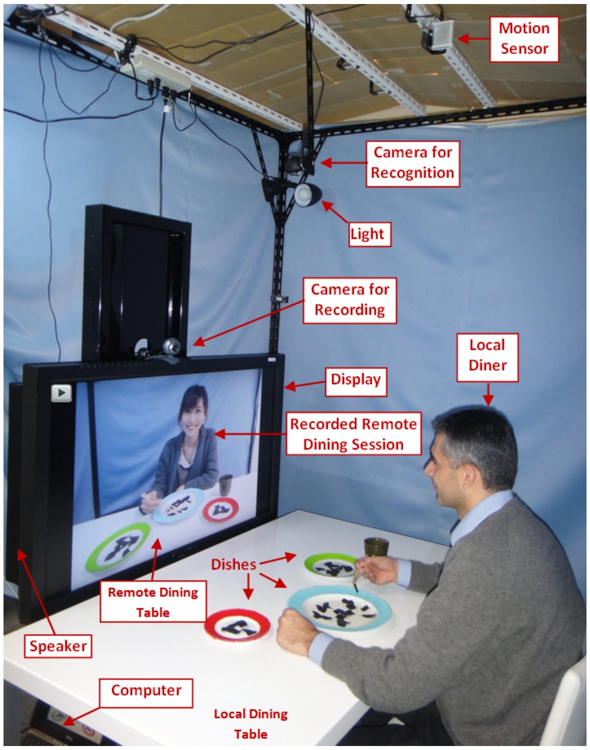

Figure 3. System's physical workspace.

User detection: The system utilizes a generic motion sensor to detect any human motions within a specified interest volume. The human motion sensor "AT Watch NET IR_mini" is used for this purpose. This sensor is fixed above the dining table at a height of 230 cm. This way, the sensor is able to detect any motion within a ground radius of 230 cm.

Food detection: This module detects any food placed on the dining table, in two steps; first step: detecting dishes on the dining table; second step: detecting the total amount of food in the detected dishes. In this study, we assumed that all the food would be served 'upfront', with the food placed directly onto the plates and brought to the table at once [7].

- Dish detection: this is performed using shape and color recognition. A previously devised method [11] is used to detect the dishes. A USB camera is fixed above the dining table at a height of 200 cm, to exclusively capture the dining table from above. At this stage of construction, we considered white circular dishes only. In addition, we used dishes with solid colored rims to easily distinguish the dishes. The free OpenCV library is utilized to perform this task, and to determine the dishes' position on the dining table.

- Amount of food detection: this is achieved using image processing techniques. A previously devised method [11] is used to estimate the remaining food from a 2D image of the dishes. At the beginning of the dining session, all the detected amount of food in all the dishes is summed together to produce a total initial amount of food. Periodically, this detection is repeated to determine the amount of consumed food compared to the total initial amount. The process ends when the user is no longer present at the dining table and/or the remaining amount of food becomes zero.

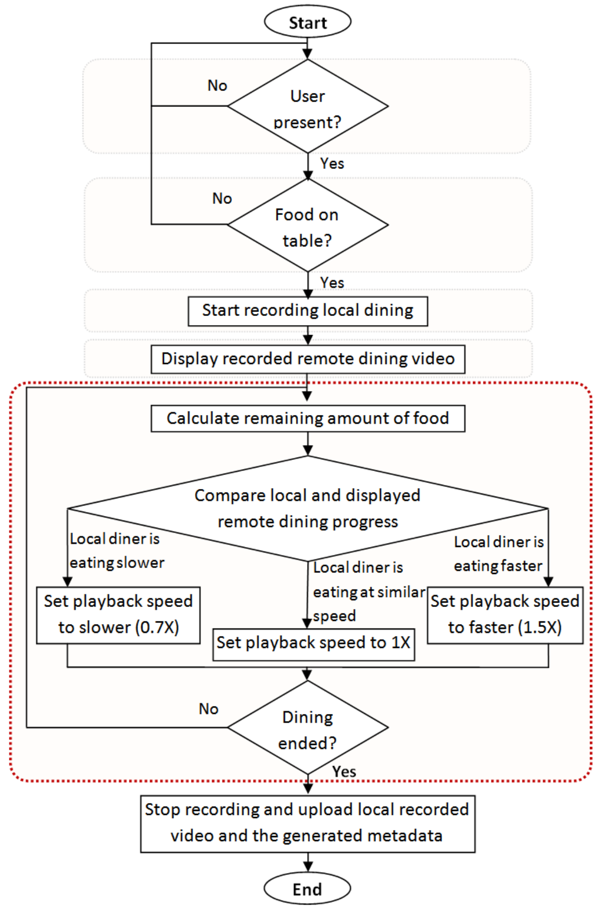

Figure 4. System's procedure flowchart.

Video recording: Recording of a dining session starts when actual dining activity is detected. This includes a person sitting at the dining table (User detection), and food placed on that table (Food detection). Accordingly, the beginning of a dining session is indicated by the diner moving his/her hand over one of the dishes. This is achieved using hand detection by the background subtraction method. A previously devised method [11] is used to detect the user's hands on the dining table.

As soon as a dining session starts, a USB camera placed over the large display in front of the person starts capturing images, and a pin microphone starts capturing sounds. The Microsoft DirectShow application programming interface (API) is used to create a media file (.avi) from the captured images and sounds. During the dining session, periodic dining progress information is collected and saved to a designated metadata file. This information includes a time stamp, a meal progress percentage, and whether or not user motion was detected at that time.

At the end of each dining session, the recorded video and the metadata file are uploaded to cloud storage in a designated user's folder. Figure 5 shows the internal folder structure of the cloud storage. In this structure, each user has a folder containing all the recorded dining sessions. Each dining session consists of a single video file and a designated metadata file. Dining sessions seen by other users are automatically archived in the user's old dining sessions folder.

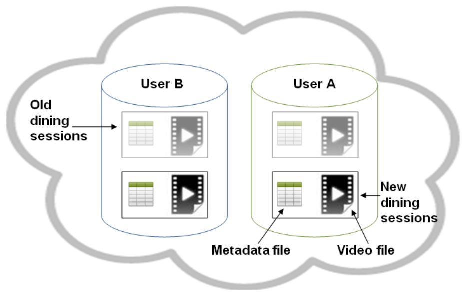

Figure 5. Internal folder structure of the cloud storage.

Video streaming and controlling: As soon as an actual dining session starts, the most recent recorded remote dining video is streamed and displayed on a large display in front of the local diner. The libVLC API[3] is used to play and control the streamed video. In order to ensure that the displayed recorded remote dining progress status matches the local dining progress status, the displayed video playback speed is controlled. The system accesses information from the metadata file and compares it to the current detected local dining progress. Based on this comparison (remote percentage eating progress minus local percentage eating progress), the system sets the displayed video playback speed. Through experimentation, we found that diners can watch a video with adjusted playback speed in a range from 0.7X to 1.5X without losing any information or sensing any distortion. Thus, the playback speed adaptation method is based on this range. If the percentage DDP is less than a specific amount, the playback-speed is set to slow-speed mode (0.7X). If the percentage DDP is greater than a specific amount, the playback speed is set to fast-speed mode (1.5X). Otherwise, the playback speed is set to normal-speed mode (1X).

THE VALIDATION OF THE ADAPTATION METHOD

In this section, the experimental validation of the KIZUNA's adaptation method is discussed. The main objective of the experiment was to determine whether using KIZUNA's adaptation method would enhance a given diner's communication and sense of presence of the remote diner, in comparison with conventional time-shifted tele-dining. Synchronizing the recorded remote dining activity with the local dining progress is expected to enhance these factors. The participants' feelings and feedback were assessed by means of a questionnaire filled out after each tele-dining experimental session.

Recording Remote Dining Video Messages

In real-life social dining, a person usually enjoys the company of familiar people, such as family members, friends or co-workers. This way, the person feels more relaxed, communicating easily and without constraint [6]. In the validation experiment, we achieved this by inviting potential participants to dine while watching a recording of another familiar person (or 'actor'). This limited the pool of participants. To extend the pool of participants, we recorded three dining sessions, each with a different actor, in a different language (Japanese, Chinese and Arabic). To achieve consistent eating progress among actors, we recruited three adult male graduate students ($A1$: Japanese, $A2$: Chinese, and $A3$: Arab). The actors were instructed to dine at a normal, relaxed speed, and provided with a list of general questions and comments to be delivered in their native languages, at specified times, while eating. Table 1 shows the list of these general questions and comments, and the time at which they are to be delivered. The questions and comments were collected and extracted from a typical basic dialog between two persons conversing at a distance. The time between each question and comment was determined based on the expected average length of reply.

Begin	Q) Hello, how are you today?
00:45 (mm:ss)	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. List of general questions and comments.

In the recording sessions, we used 400 g of curried rice and a glass of juice as the meal. This amount is considered sufficient for an adult person. We chose curried rice as a widely enjoyed meal. Moreover, this kind of food is often eaten with a spoon, while other kinds of food might involve the use of different utensils, such as forks or chopsticks, and this, we found through experimentation, might affect the dining progress. Analysis of the actor's recorded video showed that this amount of food (400 g) was consumed in nine minutes on average.

Dining Conditions

In the experiment, we considered the following two time-shifted tele-dining conditions:

- **Normal condition**: participant dines while watching a recorded video of a familiar person eating sometime earlier.
- **KIZUNA condition**: same as normal condition, but with the video playback speed adjusted according to the DDP.

[3]http://www.videolan.org/

Using the same amount of food in the experiment as in the recording sessions caused small variations in terms of the DDP. Furthermore, 400 g may represent either a large or small amount of food to different participants, which may cause them to feel uncomfortable while dining. Accordingly, after surveying meals sold in the market, as well as participant preferences, we decided to use the following meal amounts in the experiment, according to each participant's preference:

- Small meal: 300 g of curried rice (25% less than the amount used in the recording sessions)
- Large meal: 500 g of curried rice (25% more than the amount used in the recording sessions)

We performed a preliminarily dining experiment, and ran a simulation, to investigate the playback-speed adaptation method, and to determine the adaptation values. Figure 6 shows one participant's eating progress (large meal) compared to that of the displayed remote diner. The overall DDP was around 3 minutes, with an average of 14% DDP over the length of the meal.

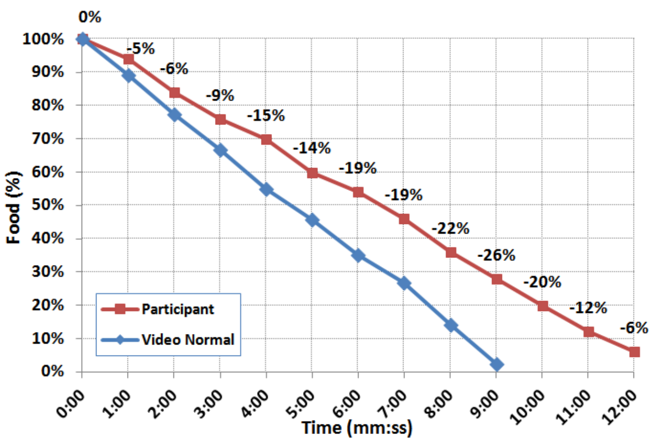

Figure 6. A participant's eating progress compared to that of the recorded remote diner. (The numbers above the chart data point indicate the percentage difference in DDP between the sessions.)

To achieve fine synchronization using the specified meal amounts, the following adaptation values were determined:

- More than -5% DDP: the local diner is considered to be eating slower than the displayed video, and the video playback speed should be changed to slow-speed mode (0.7X).
- More than 5% DDP: the local diner is considered to be eating faster than the displayed video, and the video playback speed should be changed to fast-speed mode (1.5X).
- Between -5% and 5% DDP: the local diner is considered to be eating at a normal speed compared to the remote diner, and the video playback speed should be set at normal speed (1X).

Figure 7 shows a participant's eating progress compared to the displayed remote diner progress under the KIZUNA condition. After 2 minutes of dining, the DDP exceeded the limit (-5%), so the playback speed was changed to slow-speed mode (0.7X). At the 6th minute, the DDP decreased to -2%, so the playback speed was changed to normal-speed mode (1X). At the 7th minute, the DDP increased to -5%, so the playback speed was again changed to slow-speed mode (0.7X). Under this condition, the overall DDP was reduced to only 30 seconds, with an average of -4% DDP over the length of the meal, because of the playback-speed adaptation.

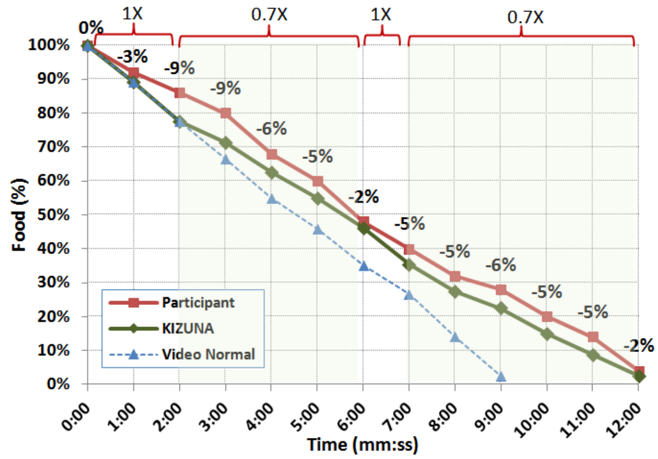

Figure 7. A participant's eating progress compared to that of the recorded remote diner, under the KIZUNA condition.

Setup

The validation experiment for the proposed adaptation dining method took place in a custom-built time-shifted tele-dining booth. The booth was enclosed by a curtain to isolate it from the lab environment. Figure 8 shows the booth's internal setup. In the tele-dining site, we used a USB camera, placed over the display, to record the participant's facial expressions and hand gestures. A small flat cooking scale was placed under the plate to accurately measure the food amount. A second USB camera was used to capture and record the cooking scale reading.

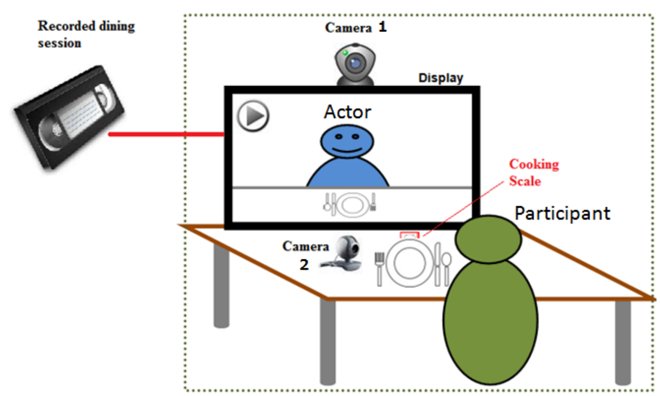

Figure 8. Tele-dining site setup.

For KIZUNA validation, the assessment of the participant's eating progress was performed by employing the WoZ system simulation technique. The second USB camera was connected to a display placed outside the tele-dining site. The researcher monitored the DDP and periodically altered the displayed video playback speed.

Group 1					Group 2				
Particpant	Sex	Video #	Bond	Meal (g)	Particpant	Sex	Video #	Bond	Meal (g)
d1	M	***V2***	Roommate	300	***d12***	F	***V3***	Friend	500
d2	M	***V1***	Friend	500	***d13***	F	***V3***	Spouse	300
d3	F	***V2***	Friend	300	***d14***	M	***V1***	Friend	500
d4	M	***V1***	Friend	500	***d15***	F	***V1***	Friend	300
d5	M	***V1***	Friend	300	***d16***	F	***V2***	Friend	300
d6	F	***V1***	Friend	300	***d17***	F	***V2***	G-Friend	500
d7	F	***V3***	Friend	500	***d18***	M	***V2***	Friend	500
d8	F	***V3***	Friend	300	***d19***	M	***V1***	Friend	300
d9	F	***V1***	G-Friend	500	***d20***	F	***V1***	Friend	500
d10	M	***V1***	Friend	500	***d21***	M	***V1***	Roommate	500
d11	M	***V1***	Friend	500	***d22***	F	***V2***	Friend	300

Table 2. Experimental session arrangements.

Participants

A total of 22 healthy participants took part in the experiment (12 females and 10 males). The participants' ages ranged from 21 to 32 years old (average = 24.1 years, s.d. = 2.7); most were undergraduate or graduate students; and 13 participants had had previous experience of videoconferencing systems. Most had used videoconferencing mainly to talk to remote family members and/or friends. The participants were divided into two groups: Group 1 participants ($d1 - d11$) performed under the Normal condition, while Group 2 participants ($d12 - d22$) performed under the KIZUNA condition. All the participants experienced each condition for the first time.

Participants were given their choice of the amount of food (small or large meal) they would consume in the experiment, and 10 participants chose the small meal, while the other 12 chose the large meal. Participants were also given the choice of dining at lunch or dinner time, and 17 participants chose lunch, while the other 5 chose dinner. Table 2 shows the participant-video arrangements, the diners' relationship, and the chosen meals, for the two groups; Where $V1$, $V2$, and $V3$ are recording of $A1$, $A2$, and $A3$ dining sessions respectively.

Procedure

Each tele-dining experiment began with the researcher asking the participant to sign the experiment consent form, and to complete a basic demographic survey form. Following this step, the researcher guided the participant to the tele-dining site and introduced the system environment. We asked the participant to imagine him-/herself dining and talking with a family member or friend who lives in a different time-zone. At this point, we displayed the familiar partner's static image on a display in front of the dining table. We explained to the participant that he/she would be dining while watching a previously recorded video of his/her partner. The participant was told that the partner would talk to him/her while dining and then later review the current recorded session. The participant was instructed to dine as he/she normally does. The participant was not apprised of the respective experimental dining condition (normal or KIZUNA). The researcher then left the dining booth and activated the recorded video as an indication of the beginning of the dining session. After each dining session, the participant was instructed to complete a questionnaire about the session he/she had just experienced. The dining sessions were recorded on tape. Afterward, the sessions were reviewed, and the DDP calculated for each condition.

Under the KIZUNA condition, the researcher played the role of WoZ, by monitoring the participant's dining progress and controlling the displayed video playback speed accordingly. The alterations were performed each minute, according to the adaptation values noted above, to achieve fine synchronization.

The experimental design was between-subject, and the analysis was performed on the results of the participants' first-ever experience of tele-dining (whether in Group 1 or Group 2).

Questionnaire

In the questionnaire, we asked the participants to evaluate a series of statements according to the feelings they experienced during the time-shifted dining session. The principal aim of the study was to assess participants' communication with their partner, and to determine whether synchronizing the dining sessions would affect communication or not, using a selection of straightforward questions from related questionnaires from [12][22]. The following statements were used for this purpose:

- C1) "*I wanted to talk to the partner.*"
- C2) "*I enjoyed talking with the partner while eating.*"
- C3) "*The partner's talking distracted me from my meal.*"
- C4) "*The content of the conversation was natural.*"
- C5) "*The timing of the partner's delivery was natural.*"
- C6) "*I could communicate with the partner naturally.*"

The sense of the partner's presence while tele-dining is another important aspect to consider in assessing the proposed system; and synchronizing the dining sessions is believed to affect this sense of the partner's presence. In this study, we sought to determine whether our proposed system would enhance this sense or not, using common questions from related questionnaires from [14][17][8]. To investigate the participants' sense of presence under each condition, the following statements were used:

- P1) "*I felt as if the partner and I were eating together in the same room.*"
- P2) "*I felt distant from the partner.*"
- P3) "*The partner's facial expressions were easy to recognize.*"
- P4) "*The partner's gaze direction was easy to recognize.*"
- P5) "*I was able to make eye-contact with the partner.*"
- P6) "*The partner's gestures were easy to recognize.*"

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.

Statistical Results

Figure 9 shows the average results of participants' communication-related feedback on their dining experience, under the two conditions. Comparison was done using a between-groups t-test, to examine whether the means under the two conditions differed significantly from one another. We found significant differences in participants' feelings of distraction caused by the partner talking (C3: t(20) = 2.62, $p < 0.05$) and in the timing of the partner's delivery (C5: t(20) = 3.66, $p < 0.01$); while no significant differences were found in participants' willingness to talk with the partner (C1: t(20) = -0.68), enjoyment of talking with the partner while eating (C2: t(20) = -1.6), content of the conversation (C4: t(20) = -1.4), or ability to communicate with the partner naturally (C6: t(20) = -0.81).

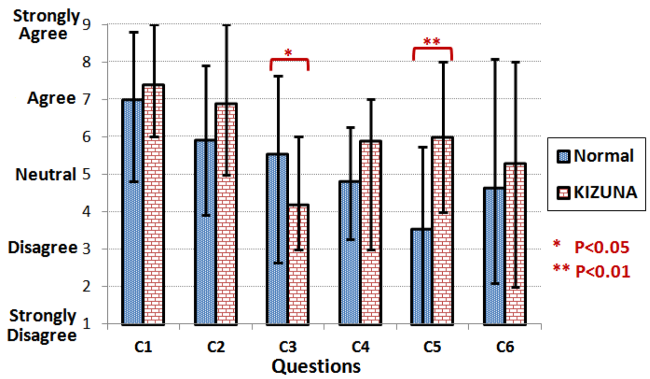

Figure 9. Participants' average communication-behavior results.

Figure 10 shows the average results of participants' felt sense of the partner's presence while dining, under the two conditions. We found significant differences in participants' sense of the partner's presence as if in the same room (P1: t(20) = -3.48, $p < 0.01$), and participants' felt distance from the partner (P2: t(20) = 2.07, $p < 0.1$); while no significant differences were found in recognizing the partner's facial expression (P3: t(20) = -0.11), the partner's gaze direction (P4: t(20) = -0.61), or the partner's gestures (P6: t(20) = -0.51), or in the ability to make eye-contact with the partner (P5: t(20) = 0.28).

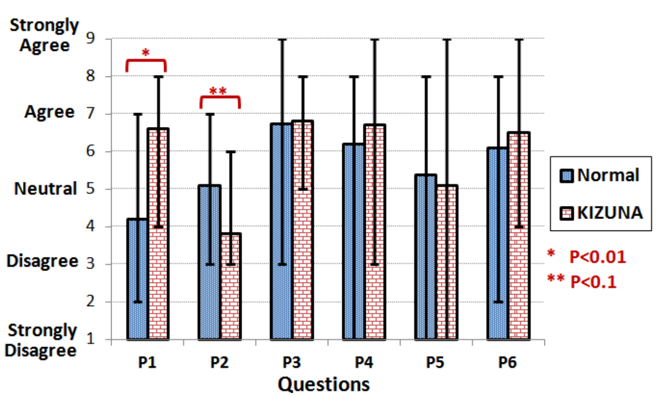

Figure 10. Participants' average sense of presence results.

The exact deployment of the $V1$ to $V3$ videos was not identical in both conditions, although all the actors were instructed to behave in the same manner, according to the scenario in Table 1: the $V1$ video was used 7 times, the $V2$ video 2 times and the $V3$ video 2 times under the Normal condition; while the $V1$ video was used 5 times, the $V2$ video 4 times and the $V3$ video 2 times under the KIZUNA condition. To determine whether the videos had any effect on the ratings, correlations between the videos and the ratings were examined for $P1$, $P2$, $C3$, and $C5$, which indicated the differences between the conditions. We found a weak correlation in Normal $P2$ (ρ = 0.41) and no correlation for the others. However, this Normal $P2$ used more $V1$ videos than under the KIZUNA condition, and the $V1$ video actually had relatively lower ratings on average. This suggests that the marginal significant difference in $P2$ is not caused by the video effect.

The proposed adaptation algorithm meant to synchronize the participant's and partner's dining progress by keeping the DDP within a specific range ($\pm 5\%$). This in turn reduced the overall DDP. For the Normal sessions, the average DDP was 117 seconds (s.d. = 48.8), while for the KIZUNA sessions, the average DDP was only 19 seconds (s.d. = 15.9), and this suggests a significant difference in the DDP (t(20) = 6.07, $p < 0.001$).

DISCUSSION

The questionnaire results revealed that the proposed time-shifted tele-dining method improved participants' communication while dining, in comparison with the Normal mode. This result supports our assumption that synchronizing the dining sessions (local and remote recorded dining sessions) would increase the realism of virtual social dining. In particular, participants found that the partner's talk in KIZUNA dining was less distracting than in Normal dining (Fig. 9 - C3). This may be a result of the close alignment of the dining progress. Moreover, participants found the partner's conversational delivery timing more natural under KIZUNA dining conditions than under Normal dining conditions (Fig. 9 - C5).

Since a sense of the remote participant's presence affects the local diner's communication, we tried to enhance this sense of presence by depicting the latter's life-sized image on a large display [10][13][19], and by matching the furniture [4] of the local and remote dining environments. The large display was placed in front of the local dining table, at the same level; hence the remote dining table appeared as if it was an extension of the local dining table (Fig. 1). We also assumed that synchronizing both dining sessions would further enhance the sense of presence. The questionnaire results support our assumption, confirming that participants felt as if the partner and they were eating together in the same room during KIZUNA dining (Fig. 10 - P1), and that participants felt less distance from the partner during the KIZUNA dining sessions (Fig. 10 - P2).

Though participants were aware that they were watching a recorded video, we noticed that some communicated more naturally with their partners during KIZUNA dining sessions, and many asked more questions in responding to the recorded video (with an average of 1.02 questions asked, compared to 0 in the Normal dining sessions). Many participants noted that it felt strange not to receive a direct response to their questions. This reaction is very similar to the "uncanny valley" telepresence reaction caused by humanoid robots that are highly similar to humans [15]. In the experiment, the remote person's life-sized representation, combined with the matching furniture and dining utensils, misled local viewers into trying to interact with the remote person in real time.

The proposed system achieved better communication and sense of presence by merely controlling the playback speed of the large displayed video. Our proposed algorithm had three playback-speed modes (normal, fast and slow). The fast and slow modes were carefully chosen so that viewers would not sense any distortion in the picture or sound. In this regard, we asked participants to judge their partner's dining speed. Despite the varying playback-speed mode, most participants said that the partner's dining speed was normal in general. However, a few participants noted that they felt the partner's motion was slightly slower in some sections, and attributed this to a problem with the video player. No comments were made, however, regarding the fast mode, which may indicate that this mode was acceptable.

It is natural for a person to look at their dinner partner while dining, and to observe their dining progress. In this regard, some participants noted that they adopted the strategy of watching their partner's plate, which gave them an indication of their partner's dining progress, by which they estimated how much time remained for the dining session. We also noticed a common behavior among participants, to echo their partner's actions, most noticeably the action of drinking. Whenever the partner performed a drinking action, the local participant performed a similar action, either synchronously or once he/she had finished their current action. We found no difference between the two conditions in this regard.

The controlled dialog during the experimental dining sessions was meant to be general and uniform for all participants. The dialog consisted of a series of questions and comments, delivered at specified times based on the dining progress. The preliminary conversational analysis showed that almost all the partner's questions were answered by participants, though only half of the partner's comments were responded to, usually in the form of a new question or agreement (or disagreement) with the respective comment. And again, no differences were found between the two conditions. When we reviewed the recorded experimental dining sessions, we found that the dialog was typically natural, especially when the partner's follow-up question or comment was a natural reply to the participant's utterance. This tended to engage participants more in the dialog, as noted by one participant: "*I was surprised when the partner answered my question immediately, as if we were talking face-to-face.*" However, some participants found the pre-set question/answer dialog style abnormal, preferring a dialogue style based on a story or the latest news.

As with any study, there are limitations on our ability to generalize from the achieved results. For example, the validation experiment was performed in a lab environment and the number of participants was quite small (3 actors and 22 participants). Additionally, the experimental setup assumed that all the food would be placed on the dining table beforehand and served on one plate only. In addition, our current system is designed for two diners; in the future, we will study how KIZUNA usage would scale if more than two diners wished to dine together. Moreover, there were other ecological shortcomings, such as participants' age, location, bond, nationality, and culture. Our study was limited to young student participants and actors (from 21 to 32 years old), and the participants and actors were located in the same city, which meant that they see each other frequently. The relationship between the participants and actors was mainly one of friendship (Table 2). We would also like to validate and extend the current findings with qualitative studies, such as longitudinal observations and investigations of dining behavior.

All these concerns will be taken into consideration when performing the actual system evaluation, once the system is fully implemented. The completed system will be deployed in participants' homes and the participant sample will consist of actual family members living apart. In addition, the system will be deployed for a considerable length of time, to investigate the ongoing communication/dining exchanges between local and remote participants.

CONCLUSION

In this paper, the design and validation of a proposed time-shifted tele-dining system is presented. This KIZUNA system is meant to enhance communication and the sense of connectedness between family members and friends who live apart in different time-zones. The system involves recording the remote person's dining sessions and then displaying these sessions for the local person when he/she starts eating, and vice versa. Based on the difference in dining progress, the system adapts the playback speed of the displayed remote recorded dining session.

A validation experiment was conducted to thoroughly investigate the proposed adaptation method, and the results revealed that the KIZUNA system effectively enhanced communication and the sense of the other person's presence, by synchronizing the dining sessions.

In future, we plan to integrate and extend the system to support time-shifted group tele-dining as well, which would benefit large families where more than one person is living apart. At this stage, we implemented a simple means of adapting the video playback speed by comparing the local and recorded dining progress. This can be further enhanced by analyzing the video content and locating 'rich' sections in the recorded dining sessions; for example, those that include an abundance of conversation and/or dining actions. Finally, the method assumed discrete levels of video playback speeds (slow, normal and fast), and this may affect the diner's viewing experience.

ACKNOWLEDGMENTS

This research was partially supported by the JSPS Grant-in-Aid for scientific research 22500104, 23500158, the Telecommunications Advancement Foundation, and the Research Projects of Graduate School of Library, Information and Media Studies, University of Tsukuba.

REFERENCES

1. Cao, X., Sellen, A., Brush, A. B., Kirk, D., Edge, D., and Ding, X. Understanding family communication across time zones. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, CSCW '10 (2010), 155–158.
2. Foley-Fisher, Z., Tsao, V., Wang, J., and Fels, S. Netpot: easy meal enjoyment for distant diners. In *Proceedings of the 9th international conference on Entertainment computing*, ICEC'10 (2010), 446–448.
3. Fukuda, S., Ed. *Emotional Engineering: Service Development*, 1 ed. Springer, 2010.
4. Gibbs, S. J., Arapis, C., and Breiteneder, C. J. Teleport towards immersive copresence. *Multimedia Systems 7* (May 1999), 214–221.
5. Heather, K. Identifying nutrition problems in senior patients. *Geriatrics and Aging 7* (2004), 62–65.
6. Hetherington, M., Anderson, A., Norton, G., and Newson, L. Situational effects on meal intake: A comparison of eating alone and eating with others. *Physiology and Behavior 88* (2006), 498–505.
7. Hupfeld, A., and Rodden, T. Laying the table for hci: uncovering ecologies of domestic food consumption. In *Proceedings of the 2012 conference on Human Factors in Computing Systems*, CHI'12 (2012), 119–128.
8. Ichikawa, Y., Okada, K., Jeong, G., Tanaka, S., and Matsushita, Y. Majic videoconferencing system: experiments, evaluation and improvement. In *Proceedings of the 4th conference on European Conference on Computer-Supported Cooperative Work*, ECSCW'95 (1995), 279–292.
9. Inkpen, K., Du, H., Roseway, A., Hoff, A., and Johns, P. Video kids: augmenting close friendships with asynchronous video conversations in videopal. In *Proceedings of the 2012 conference on Human Factors in Computing Systems*, CHI'12 (2012), 2387–2396.
10. Inoue, T. Mixed reality meeting system enabling user to keep and share interpersonal distance in the real world. *Journal of Information Processing Society of Japan 50*, 1 (2009), 246–253.
11. Inoue, T. Adaptive tabletop dish recommendation by the real-time recognition of dining status. In *Proceedings of the fifth International Workshop on Informatics (IWIN'11)* (2011), 174–180.
12. Inoue, T., Okada, K., and Matsushita, Y. Integration of face-to-face and video-mediated meetings: Hermes. In *Proceedings of the international ACM SIGGROUP conference on Supporting group work: the integration challenge*, GROUP '97 (1997), 405–414.
13. Ishida, T., Sakuraba, A., and Shibata, Y. Proposal of high realistic sensation system using the large scale tiled display environment. In *Proceedings of Network-Based Information Systems (NBiS)* (2011), 444–449.
14. 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.
15. Mori, M. Bukimi no tani the uncanny valley. *Energy 7* (1970), 3335.
16. Nakagawa, M., Nagatsuka, M., Nishiyama, M., and Yoshida, Y. The function and the possibility of kyoshoku (eating together). *HortResearch*, 64 (2010), 55–65.
17. Nakanishi, H., Kato, K., and Ishiguro, H. Zoom cameras and movable displays enhance social telepresence. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11 (2011), 63–72.
18. Ocker, R., and Yaverbaum, G. Asynchronous computer-mediated communication versus face-to-face collaboration: Results on student learning, quality and satisfaction. *Group Decision and Negotiation 8*, 5 (1999), 427–440.
19. Okada, K.-I., Maeda, F., Ichikawa, Y., and Matsushita, Y. Multiparty videoconferencing at virtual social distance: Majic design. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, CSCW '94 (1994), 385–393.
20. Romero, N., Markopoulos, P., Baren, J., Ruyter, B., Ijsselsteijn, W., and Farshchian, B. Connecting the family with awareness systems. *Personal Ubiquitous Comput. 11*, 4 (2007), 299–312.
21. Sellaeg, K., and Chapman, G. E. Masculinity and food ideals of men who live alone. *Appetite 51*, 1 (2008), 120–128.
22. Sellen, A. J. Speech patterns in video-mediated conversations. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '92 (1992), 49–59.
23. Tang, J., Marlow, J., Hoff, A., Roseway, A., Inkpen, K., Zhao, C., and Cao, X. Time travel proxy: using lightweight video recordings to create asynchronous, interactive meetings. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, CHI'12 (2012), 3111–3120.
24. Tsujita, H., Yarosh, S., and Abowd, G. D. Cu-later: a communication system considering time difference. In *Proceedings of the 12th ACM international conference adjunct papers on Ubiquitous computing*, Ubicomp '10 (2010), 435–436.
25. Wei, J., Wang, X., Peiris, R. L., Choi, Y., Martinez, X. R., Tache, R., Koh, J. T. K. V., Halupka, V., and Cheok, A. D. Codine: an interactive multi-sensory system for remote dining. In *Proceedings of the 13th international conference on Ubiquitous computing*, UbiComp '11 (2011), 21–30.
26. Zuckerman, O., and Maes, P. Awareness system for children in distributed families. In *Proceedings of the Conference on Interaction Design and Children (IDC)*, ACM Press (2005).