Development of KIZUNA System to Support Time-Shifted Co-Dining

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 is an essential activity in our lives. During the mealtimes, family members get a chance to talk and share experiences. However, sometimes not everyone can join the meal either because they are busy or maybe because they lives apart. In recent years, the wide variety of videoconferencing technologies has broadened the ways people communicate with each other. But most of these technologies require that both sides are available at the same time to have communication. In this research we propose a time-shifted tele-dining system "KIZUNA" for family members to enjoy a meal together even when they are unable to physically meet together. A person can enjoy a meal while watching a pre-recorded video of a remote family member's dining. The system adapts the displayed video's playback speed according to the local diner's eating behavior. This is likely to enhance the communication and increase the enjoyment while dining.

1. Introduction

"*The family that eats together stays together*" [1].

Food plays vital role in our lives. A person cannot stay alive without food. After breathing, eating is what we do most frequently during a lifetime. Eating (also known as consuming) is defined as the ingestion of food to provide for all organisms their nutritional or medicinal needs, particularly for energy and growth [2]. Almost every house has a kitchen, where food is prepared, and a dining room, where family members sit together to eat the food. Beyond simply a place to eat, the dining room is a place where family members meet and talk. During meals, families have the chance to catch up on what has happened in each other's live, and to strengthen the bonds that hold the family together. For example, during mealtime, parents have the opportunity to find out about their children's day at school [3]. In this regards, a study by Larson et al. revealed that 80% of parents viewed family dinners as important, and 79% of teens considered eating family meals to be among their top-rated family activities [1].

Family members or close friends often enjoy meals together. In the past, family members used to meet almost daily to eat and spend social time together. On the other hand, recently, and because of the busy lifestyles, having a meal with family becomes very difficult. This is particularly true as older people live independently; younger people live alone and working people travel more and work from afar [4]. In these cases, some of family members have to eat alone. Such situation might affect people lives and makes them feel lonely and unhappy. Sadly enough, this is a scenario that faces countless families all around the world as children move out and away.

A variety of inexpensive videoconferencing technologies offers a decent solution for distant family members to maintain excellent sense of social connectedness, which is defined as "*positive emotional appraisal, characterized by a feeling of staying in touch within ongoing social relationships*" [5].

When family members are located far away from each other this often means that they are also living in different time-zones. The different time-zones are likely to pose more difficulties for communication between distance family members. The main difficulty came from the misalignment of daily schedules between the two parties of communication [6]. This clearly appears at dining time. For example, if a person was located in Japan (UTC+9) and his/her family were located in a pacific region in Canada (UTC-8), this means that when this person having a dinner at 7pm, his/her family were still sleeping (3am).

In this paper, we propose a solution "KIZUNA" to overcome the eating alone problem when only one side is available for dining. The solution is mainly based on time-shifted communication by means of transmitting recorded video messages. This involves recoding remote person's dining sessions and displays these sessions to the local person when he/she starts eating and vice versa.

The system is capable of recognizing local person's eating behavior and dining progress. Based on this information, the system adapts the playback speed of displayed remote dining session. More specifically, the system can recognize when a dining session starts. A

CollabTech 2012 , August 27-29, 2012, Hokkaido, Japan.

recorded remote dining session is then streamed and displayed on a large display placed in front of the local person. If the local person eats faster than the person in the video, the video is automatically controlled to play faster by the system, and vice versa. This solution is likely to improve the person's dining experience and increase the communication and the sense of connectedness with his/her remote family members or close friends.

The following section surveys related works. Next section introduces the system proposal and design specifications. This is followed by the system implementation. The last section ends up with the conclusion and the future work.

2. Related works

Many research fields investigated ways to utilize the information and communication technology in minimizing the gap between people living apart. In this section, we briefly overview some works related to people dining together despite their being located in remote places and/or in different time-zones.

2.1. Synchronous dining

Family members and close friends tend to have meals together, but the lack of time or being far makes this hard to 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 get together virtually [7]. This prototype was essentially a videoconferencing system targeted people with limited knowledge in technology e.g. elderly people. Thus, Accenture made it highly automated and easy to operate. The system monitored 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.

An advertising agency Wieden+Kennedy's Amsterdam office produced a website called Virtual Holiday Dinner. This website was implemented for scattered friends and family to have a dinner party of up to five people via Skype™ [8]. Guests can call into the dinner, and their faces are shown on the displays that are the heads of the dolls physically sitting around a dining table. The dolls were equipped with video cameras fitted with facial tracking software, so each guest can look around the dining table through the doll by moving his/her head.

The sense of coexistence among family members is likely to be affected when living apart. To maintain a high sense of coexistence among family members, a system called CoDine was presented [9]. This system is a dining table embedded with interactive subsystems that augment and transport 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 transportation of edible messages.

Dining together is also an enjoyable experience. In some cultures, meals are consumed in groups, usually family members or close friends. Examples of popular group meals are the Chinese Hot-Pot, the Japanese Nabe, the Arabic Mansaf, etc. In this regards, the remote group meal communication was investigated to make a meal more enjoyable [10]. In this study of Chinese Hot-Pot meal, three factors that are essential for group meal experience were identified; interacting as a group with food, a central shared hot-pot, and a feeling that others are nearby.

All of this previous research primarily focused on synchronous dining among a group of people living apart, assuming that all of them are available at the same time. Our research, in contrast, is focusing in the cases when only one side is available for dining.

2.2. Asynchronous dining

When both sides are available at the same time, dining together can be achieved by a variety of videoconferencing systems. However when only one side is available, things become more difficult. In this situation, asynchronous collaboration tools can provide a solution in general. Because users can receive the information they want when it's most convenient for them, those tools have been widely used in work settings. But at the same time, they can lack a sense of immediacy. For example in distant learning domain, Ocker et al. research 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 [11]. However, the students were less satisfied with asynchronous learning experience, both in terms of the group interaction process and the quality of group discussion. This suggests that asynchronous interaction is satisfactory for some requirements but not for some other requirements, presumably affective satisfaction or social connectedness.

In the dining domain, Tsujita et al. proposed a system "CU-Later" to be used in time-shifted dining [12]. This system plays a recorded video of remote dining after a specific time shift when the local user is in front of the display which is placed on a dining table, so the local user can watch the video automatically when he/she is eating. This system records the local user's session as well when the video is played, so similarly the remote user can watch the local user's session later on. It is like a video mail exchange with automatic playback and recording.

We think communication when dining together means more than just information exchange but affective satisfaction or social connectedness, namely the type of

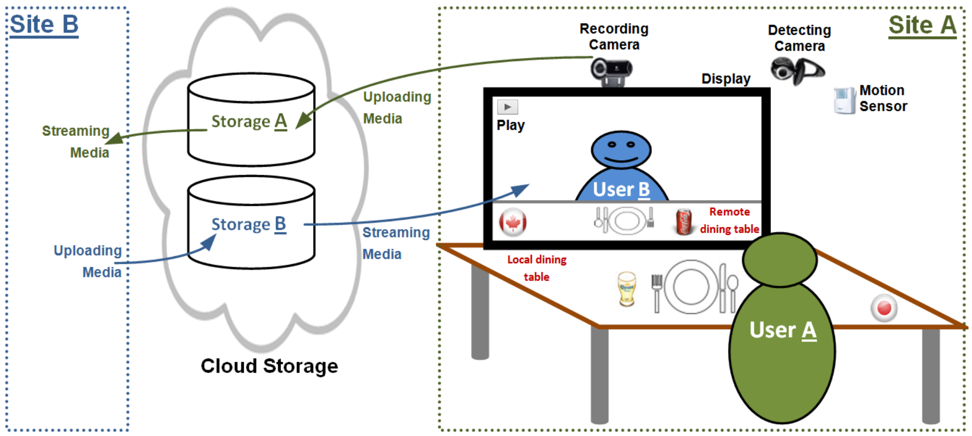

Figure 1. KIZUNA system scenario.

communication called consummatory communication. To support this type of communication, the sense of co-presence may be crucial. Thus our proposed system is also for this situation and is exploration of video exchange, but is more focused on the social connectedness between the remote users. We think that synchronizing dining activities of both sides is significant to enhance the users' dining experience.

3. System proposal and design

3.1. System proposal

Figure 1 shows a conceptual illustration of the KIZUNA (means "bond" in Japanese) system scenario. In this scenario two persons, local user A and remote user B, are virtually dining together although they are physically apart and/or are in different times.

The system automatically starts working when the local person A gets seated and some dishes are placed on the dining table. This is supposed to increase the system usage portability especially for people with limited knowledge of technology. The system starts playing the video of remote user B when he was eating sometime earlier by a large display placed in front of a local dining table in life-size scale. It also starts recording the video of the local user A dining.

This automatic playback and recording of videos is to provide the feelings of co-dining for the users in actually different times and/or different places. We think that merely watching the other user's video may not arise the feeling of co-dining. We think that synchronization of both dining activities, not only the start time but also the process, is significant. Specifically this synchronization can be realized by controlling the video playback speed to match the local user A's dining status. This way, the local user is expected to enjoy the company of the remote user until he complete his/her meal. Also in this way, we assure that the local user can see the entire remote dining session while he is eating. This is likely to enhance the dining experience.

3.2. System design

Based on the system proposal, we defined the following set of design requirements to implement the KIZUNA system:

Dining detections: the dining status and progress is necessary information to control the system. This includes the following:

- **User detection:** the system has to be highly automated and easy to operate. Information as user existence around the dining table should be considered to start the system and display the remote side's recorded dining video. This will makes the system much usable and portable especially by people with limited knowledge in technology.
- **Food detection:** detecting a user around a dining table is not enough information to assume that a dining session is going to start. We also need to detect whether there is food placed on the dining table or not.
- **Amount of food detection:** based on this information the system can judge some of the user's dining behavior. We are particularly interested in the user's eating speed.

Video manipulation: in the proposed KIZUNA system the users can communicate through exchanging recorded video message. 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 local diner's eating progress.

4. System implementation

4.1 Physical workspace

An experimental time-shifted tele-dining system is implemented based on the KIZUNA system proposal and list of design requirements. The implemented system consisted of two sites. The experimental site is equipped with the following: a dining table (135cm length by 80cm width by 70cm height), a chair, a large flat-panel display (46 inches), two USB cameras (one for recording and one for recognition), speakers, a microphone, a motion detection sensor, two spotlights, and a computer. **Fig. 2** shows the implemented system's physical workspace.

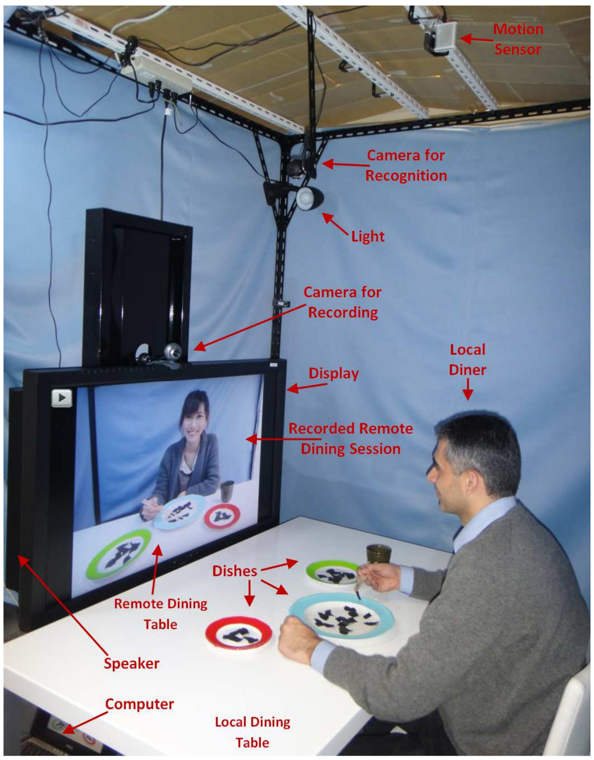

Figure 2. System's physical workspace.

4.2 Software

The system's software is implemented using Microsoft Visual C++ on windows OS. **Fig. 3** shows the procedure flowchart, which is constructed by the following major modules:

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[1]. This sensor is fixed above the dining table at a height of 230cm. This way, the sensor is able to detect any motion within a ground radius of 230cm. This module is implemented using a separate thread to check the sensor's reading every 500msec. The module returned values are: "1" when a human motion is detected and "0" when no motion is detected. **Table I** lists the used sensor's main technical specifications.

Table I. AT WATCH NET IR_mini's main technical specifications.

Detection distance		Maximum 5m
Detection range	Horizontal	100°
	Vertical	82°

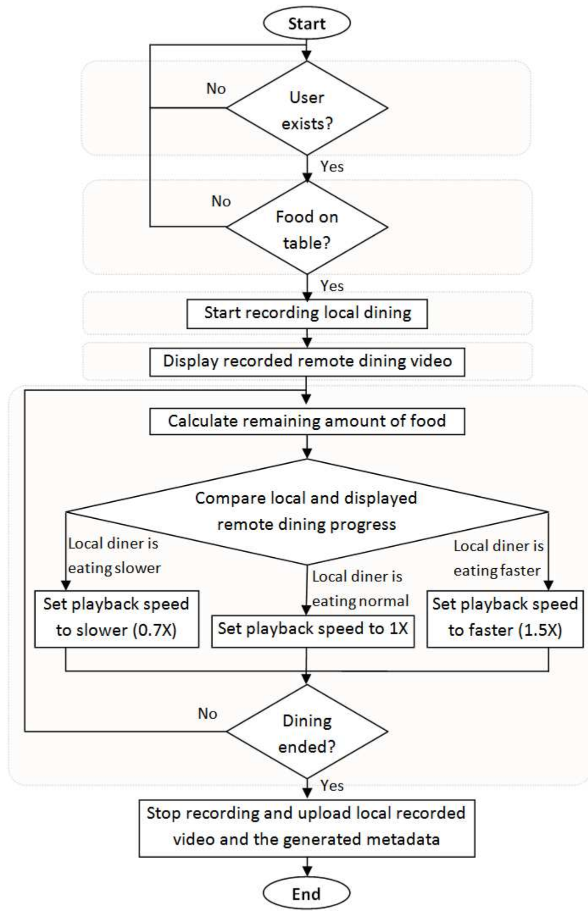

Figure 3. System's procedure flowchart.

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.

- **Dish detection:** this is performed using shape and color recognitions. A previously implemented method in [13] is used to detect dishes. A USB

[1] http://www.ntt-at.co.jp/page.jsp?id=1793&content_id=886

camera is fixed above the dining table at a height of 200cm to exclusively capture the dining table from top. At this stage of implementation, we only considered white circle dishes. Furthermore, we used dishes with unified 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 implemented method in [13] is used to estimates the remaining food from a 2D image of the dishes. At the beginning of a dining session, all the detected amount of foods in all the dishes are summed together to produce a total initial amount of food. Periodically, every 5 seconds, this detection process is repeated to find out the amount of consumed food compared with the initial total amount. This process ends when the person is no more existed around the dining table and/or the remaining amount of food becomes zero.

Video recording: recording of a dining session starts when a real dining activity is detected. This includes, a person sitting around a dining table (User detection), and food placed on that table (Food detection). Accordingly, the beginning of a dining session is determined once the person moves his/her hand over one of the dishes. This is implemented using hand detection by background subtraction method. A previously implemented method in [13] is used to detect 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) out from the captured images and sounds.

During the dining session, periodic dining progress information are collected and saved to a designated metadata file. This information includes a time stamp, a meal progress percentage (100% means all the food in the dishes were consumed), and whether a user motions was detected or not at that time (1 means user's motion was detected). **Table II** shows a sample dining metadata file.

Table II. A sample metadata file.

	A	B	C	D
1	Tag	Progress %	Time sec.	Exists
2	data0	26	0	1
3	data1	45	8	1
4	data2	74	31	1
5	data3	97	39	1

At the end of each dining session, the recorded video and the metadata file are uploaded to cloud storage under a designated user's folder. **Fig. 4** shows the internal folder structure of the cloud storage. Under this structure, each user has a folder contains all the recorded dining sessions. Each dining sessions consisted of a single video file and a designated metadata file. The dining sessions which seen by the other user is automatically archived under the user's old dining sessions folder.

Video streaming and controlling: as soon as a real dining session starts, the last pre-recorded remote dining video is streamed and displayed on a large display placed in front of the local person. The libVLC API from VideoLAN[2] is used to play and control the streamed video. In order to assure that the displayed recorded remote dining progress status match the local dining progress status, the displayed video playback speed is controlled. The system uses the information from the metadata file and compares it to the current detected local dining progress. Based on this compression (remote percentage eating progress minus local percentage eating progress), the system set the displayed video playback speed accordingly. By experimentation, we found that the person can watch a video with adjusted playback speed in the range from 0.7X to 1.5X without losing any information or feeling any distraction. Thus, the playback speed adaptation method is based on this range. If the percentage dining progress difference is less than a specific amount, the playback seeped is set to slow speed mode, i.e., 0.7X. If the percentage dining progress difference is greater than a specific amount, the playback speed is set to fast speed mode, i.e., 1.5X. Else the playback speed is set to normal speed mode, i.e., 1X.

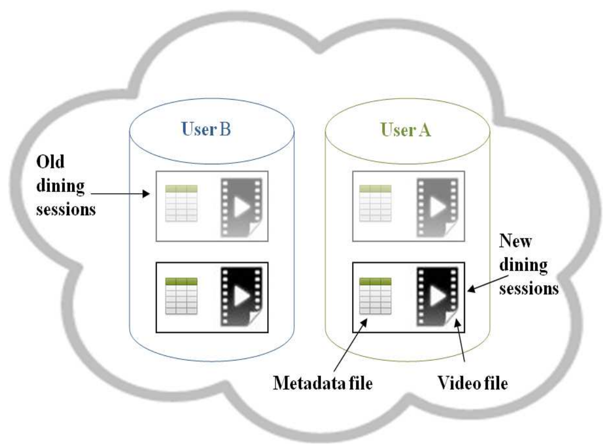

Figure 4. The internal folder structure of the cloud storage.

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

5. Conclusion and future work

In this paper, the design and the implementation of a proposed time-shifted tele-dining system is presented. This system "KIZUNA" meant to enhance the communication and the sense of connectedness between family members and close friends who live apart from each other in different time-zones. Our solution involves recoding remote person's dining sessions and displays these sessions later on to the local person when he/she starts eating and vice versa. Based on the diner's eating behavior, the system adapts the displayed remote recorded dining session's playback speed.

In the future, we are planning to conduct a system evaluation experiment. The participants' interaction will be thoroughly investigated in both directions. Also, we are planning to extend the system by supporting time-shifted group tele-dining as well. This might be necessary for big families where more than one person is living afar. Regarding the method we used to detect the amount of food, the initial assessment shows that it works fine on a limited kind of foods. In order to enhance the method and the detection accuracy, other methods will be tested out.

At this stage, we implemented a simple way of adapting the video playback speed by comparing the local dining and recorded dining progress. This can be enhanced further by also analyzing the video content and focus on the rich sections in the recorded dining sessions. For example, a rich section might be that section that includes a lot of talks and dining actions. Other adaptation technique involves video summarization by visualize a time sequence as in [14]. This technique enables viewer to quickly look into a long video message and understand the content. Finally, the method assumed discrete levels of video playback speeds, i.e. slow, normal, and fast, this might affects the person's watching experience. In a recent study, they found that a person preferred gradual rather sudden or dramatic increase of playback speed [15]. Accordingly, a gradual continuous video playback speeds will be considered in the future implementation.

Acknowledgment

The authors would like to thank Yasuhiro Ishikawa for his contribution to the development of this system. 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] R. Larson, K. Branscomb, and A. Wiley, "Family mealtime as a context of development and socialization", *New directions for child and adolescent development*, Jossey-Bass, 2006.

[2] Wikipedia: Eating. Retrieved March 25, 2012, from http://en.wikipedia.org/wiki/Eating .

[3] D. Bell and G. Valentine, "Consuming geographies: we are where we eat", *Cultural studies, Geography, Routledge*, 1997.

[4] K. Sellaeg and G. E. Chapman, "Masculinity and food ideals of men who live alone", *Appetite*, 51(1), 2008, pp.120-128.

[5] N. Romero, P. Markopoulos, J. Baren, B. Ruyter, W. Ijsselsteijn, and B. Farshchian, "Connecting the family with awareness systems", *Personal and Ubiquitous Computing*, 11(4), Apr. 2007, pp. 299-312.

[6] X. Cao, A. Sellen, A. B. Brush, D. Kirk, D. Edge, and X. Ding, "Understanding family communication across time zones", *In Proceedings of the 2010 ACM conference on Computer supported cooperative work, CSCW'10*, New York, NY, USA, 2010, pp. 155-158.

[7] Gizmodo: Virtual Meals Let You Pig Out with Distant Relatives. Retrieved March 25, 2012, from http://gizmodo.com/accenture-virtual-family-dinner/ .

[8] Virtual Holiday Dinner. Retrieved March 25, 2012, from http://www.virtualholidaydinner.com/ .

[9] J. Wei, X. Wang, R. L. Peiris, Y. Choi, X. R. Martinez, R. Tache, J. T. K. V. Koh, V. Halupka, and A. D. Cheok, "Codine: an interactive multi-sensory system for remote dining", *In Proceedings of the 13th international conference on Ubiquitous computing, UbiComp '11*, New York, NY, USA, 2011, pp. 21-30.

[10] Z. Foley-Fisher, V. Tsao, J. Wang, and S. Fels, "Netpot: easy meal enjoyment for distant diners", *In Proceedings of the 9th international conference on Entertainment computing, ICEC'10*, Berlin, Heidelberg, Springer-Verlag, 2010, pp. 446-448.

[11] R. J. Ocker and G. J. Yaverbaum, "Asynchronous computer-mediated communication versus face-to-face collaboration: Results on student learning, quality and satisfaction", *Group Decision and Negotiation*, 8(5), 1999, pp. 427-440.

[12] H. Tsujita, S. Yarosh, and G. D. Abowd, "Cu-later: a communication system considering time difference", *In Proceedings of the 12th ACM international conference adjunct papers on Ubiquitous computing, Ubicomp'10*, New York, NY, USA, 2010, pp. 435-436.

[13] T. Inoue, "Adaptive Tabletop Dish Recommendation by the Real-time Recognition of Dining Status", *In Proceedings of the fifth International Workshop on Informatics (IWIN2011)*, 2011, pp. 174-180.

[14] C. Nguyen, Y. Niu, and F. Liu, "Video summagator: an interface for video summarization and navigation", *In Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems, CHI '12*, New York, NY, USA, 2012, pp. 647-650.

[15] K.-Y. Cheng, S.-J. Luo, B.-Y. Chen, and H.-H. Chu, "Smartplayer: user-centric video fast-forwarding", *In Proceedings of the 27th international conference on Human factors in computing systems, CHI '09*, New York, NY, USA, 2009, pp. 789-798.