Estimating Human Interruptibility in the Home for Remote Communication

Yoshinao Takemae
NTT Cyber Solutions Laboratories, NTT Corporation
1-1 Hikari-no-oka, Yokosuka-Shi, Kanagawa, 239-0847 Japan
takemae.yoshinao@lab.ntt.co.jp (also with Keio University)

Takehiko Ohno
NTT Cyber Solutions Laboratories, NTT Corporation
1-1 Hikari-no-oka, Yokosuka-Shi, Kanagawa, 239-0847 Japan
ohno.takehiko@lab.ntt.co.jp

Ikuo Yoda
NTT Cyber Solutions Laboratories, NTT Corporation
1-1 Hikari-no-oka, Yokosuka-Shi, Kanagawa, 239-0847 Japan
yoda.ikuo@lab.ntt.co.jp

Shinji Ozawa
Department of Information and Computer Science, Keio University
3-14-1 Hiyoshi, Kohoku-Ku, Yokohama-Shi, Kanagawa, 223-8522 Japan
ozawa@ozawa.ics.keio.ac.jp

CHI 2006, April 22–27, 2006, Montréal, Québec, Canada.
ACM 1-59593-298-4/06/0004.

Abstract
This paper presents a method for automatically estimating human interruptibility in home environments. To make online remote communication smoother, determining if it is appropriate to interrupt the remote partner is critical. As a first step in achieving this goal, several audio-visual features, extracted from data streams provided by a camera and a microphone, are correlated to human interruptibility. Based on these features, the level of interruptibility is estimated using the trained Support Vector Regression (SVR) technique. Finally, we discuss the potential of our method based on the results of several experiments.

Keywords
Interruptibility, presence, awareness, audio-visual tracking, home, online remote communication

ACM Classification Keywords
H1.2. [MODEL AND PRINCPLES] User/Machine Systems, H4.3. [INFORMATION SYSTEMS APPLICATIONS] Communication Applications

Introduction
With the growth of recent broadband networks such as optical fiber and ADSL, online remote communication services such as the videophone and teleconferencing

are becoming as popular as telephones and cellular phones. The next-generation network must provide advanced online remote communication services for realizing smooth and efficient communication like face-to-face communications.

In face-to-face communications, humans find it easy to start communication, i.e. to interrupt someone, because we can determine the partner's situation from verbal and nonverbal cues like gaze and facial expression etc. However, in remote communication, humans cannot do the same since really effective feedback is not available until the partner accepts the call. For example, most of us have regretted calling someone at the wrong time and wished that we had known that the time was inappropriate before we called.

One means of conveying information about the intended partner’s situation is set a video camera at the partner's site. This approach, however, is unacceptable since it leaks private information. The need to ensure human privacy is paramount to gain user acceptance.

Our goal is to support remote communication so that people can start communication when it is appropriate for the partner. To this end, it is essential that the interruptibility of the partner be acquired without leaking private information. Using this information allows the user to determine the appropriate timing to contact the partner intended.

In order to advance research towards this goal, we focus here on the home and propose an automatic method that estimates the level of human interruptibility based on audio-visual tracking.

Related Works

Studies on interruptions

Although interruption was first studied in the field of classical psychology, it has also been the focus of recent researches in HCI with the view of realizing design guidelines for deciding when to interrupt people with PC tasks [1]. Hudson et al. investigated key human behaviors that might be related to human interruptibility in private offices (working environments) [2]. Based on their analysis of the results collected, they found that the presence/absence of talking was a significant indicator of the level of interruptibility. However, since they manually tagged voice and video streams, this method has not been implemented as a practical system. Minakuchi et al. proposed a method for automatically estimating the user’s interruptibility at his/her desk [3]. This method assigns one of three levels of activity according to the usage of pens, the presence of sound, keystrokes, and the movements of the mouse cursor. Experiments showed that the resulting estimations were 43.8% accurate. However, very little research has been done that focused on human interruptibility in home environments.

Tracking human activity in home environments

With recent growth of ubiquitous computing techniques, many works such as *Aware Home* [4] and *Ubiquitous Home* [5] have attempted to make human everyday life easier and richer in the home. These studies used devices such as cameras, microphones, and proximity sensors to detect human activity as robustly as possible. However, most works [6] considered only simple actions like sitting, standing etc. Very little has been achieved in detecting the more complicated attribute of human interruptibility.

Data Collection

To capture and analyze data about human interruptibility in a living room, we placed a PC-based audio-visual capture system in one Japanese living room of a family of four.

Capture System

The Windows PC was connected to a USB camera, a microphone, and a ten key keyboard. The camera had a wide-angle lens and was attached to a lighting fitting on the ceiling. The gray scale output image (Mpeg format) had a resolution of 320x240 pixels at about 25 [frame/s], as shown in Figure 1. The directional microphone was set on a wall and directed at the center table. 16-bit mono audio data was recorded at 44 [KHz]. The keyboard, used by the subjects to input the level of interruptibility, was set on a corner of the table. Recording was performed from 8am to 23pm for 5 days on 2005 Dec. (winter).

Subjects

We chose two subjects who often pick up and respond to telephone calls to the family. One was a 56-year old housewife. The other, a 22-year old daughter, had a weekday job.

Procedures

In order to capture natural data, subjects were asked to live life as usual. A vibration timer with repeat function was attached to each subject. This timer vibrated each 15 minutes. If the timer vibrated when the subject was in the living room, the subject entered her current level of interruptibility as soon as possible using a 5-point scale: 1 (lowest level of irritation with the interruption) to 5 (highest level of irritation with the interruption) and select current behavior such as eating and drinking, speaking, sleeping, reading, writing, watching TV, listening to music, cleaning a room, and folding the washing. We eliminated the data collected in the morning of the first day since it was taken to cover a training period. Consequently, we collected a total of 127 data points from the two objects.

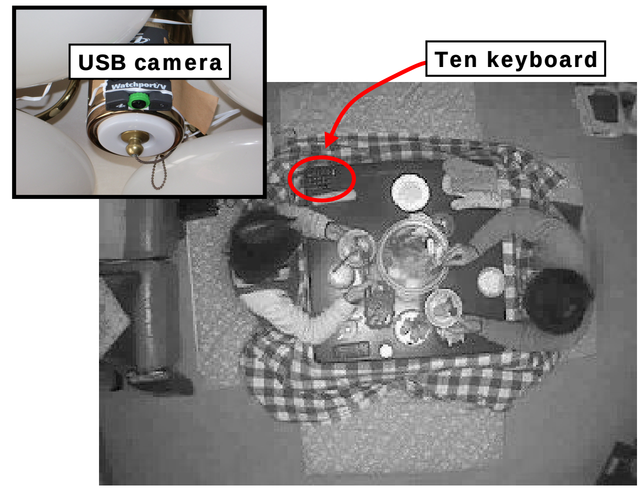

Figure1. Arrangement of USB camera and typical image.

The Data

This section characterizes the data collected.

Overall distribution of interruptibility

Figure 2 shows the overall distributions of interruptibility for the two subjects. The difference between the subjects is slight; the least (most)-interruptible level was selected 8.1 (29.3) % of the time.

Human behaviors for estimating human interruptibility

We investigated how human behavior was related to human interruptibility in a living room. The main results are as follows.

- Speaking. 89.2 % of the 56 data points indicating this activity were distributed from level 2 to 5.
- Eating and drinking. 100 % of the 6 samples indicating this activity were distributed from level 3 to 5.
- Sleeping. 100 % of the 16 samples indicating this activity were assigned level 5 (greatest irritation).

Proposed Method for Estimating Human Interruptibility

Overview

Our method consists of online three modules: data collection, feature extraction, and estimation. In data collection, a USB camera captures visual data of occupants in the room (See Figure 1). A microphone captures occupants' audio data. Feature extraction identifies several audio-visual features such as voice power and occupants' motion. Estimation uses a trained Support Vector Regression (SVR) system [7] to output the level of human interruptibility for the room targeted.

Feature extraction

To create a really practical system, we focused on several low-level features that could be extracted with relatively little delay. Due to space constraints, only the key features are described below. We calculated the average and variance of each feature over the past N [seconds] at each time of self-data entry.

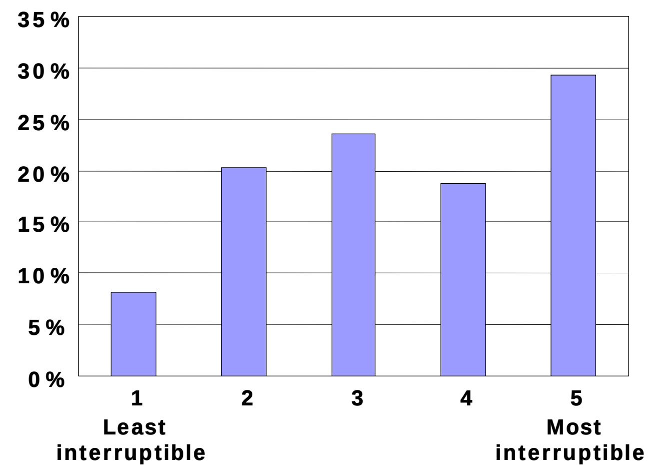

Figure 2. Overall distribution of interruptibility levels of two subjects.

Audio features

Mirroring past research [1], we found that talking with some one else in the room makes for an interruptible state. Accordingly, the logarithmic powers of recorded audio data at time t were calculated.

Visual features

(V1) Motion on the table

Based on our finding that eating and drinking trigger relatively high levels of human interruptibility, the changes of image intensity inside the table area (*hotspot R*) were extracted based on frame differences between current image and previous image using equation (1). To reduce the noise created by changes in illumination, every pixel with an absolute value of frame differences smaller than threshold T was set to zero. Figure 3 (c) shows the results. In this figure, each white pixel represents a change due to motion.

$$V1(t) = \sum_{(x,y)\in R} \left| I(x,y,t) - I(x,y,t-1) \right| \tag{1}$$

(V2) Relative change in location

Based on our finding that static behavior like sleeping yields the highest level of human interruptibility, the relative change in the center of occupant mass ($m_x(t)$, $m_y(t)$) was extracted using equation (2). Figure 3 (d) shows occupant locations as identified by a background subtraction technique and updated background images. In this figure, white areas show occupant locations.

$$V2(t) = \left| m_x(t) - m_x(t-1) \right| + \left| m_y(t) - m_y(t-1) \right| \tag{2}$$

(V3) Relation in position among occupants

From our observations we hypothesized that occupants are close to each other when engaged in the same behavior (eating and talking etc.) and cooperative work (teaching etc.), which suggests high levels of interruptibility. The radius of a smallest circle that encloses occupant masses at time *t* was calculated, as shown in Figure 3 (d). In this figure, the blue circle shows the circle enclosing the occupants.

Estimation

The level of human interruptibility was estimated using a trained Support Vector Regression (SVR) system. SVR is an optimization-based approach for solving machine learning regression problems based on the Support Vector machine (SVM), which is widely used for human modeling and pattern recognition [7]. The system was constructed using a widely available, open source library (LIBSVM) [8].

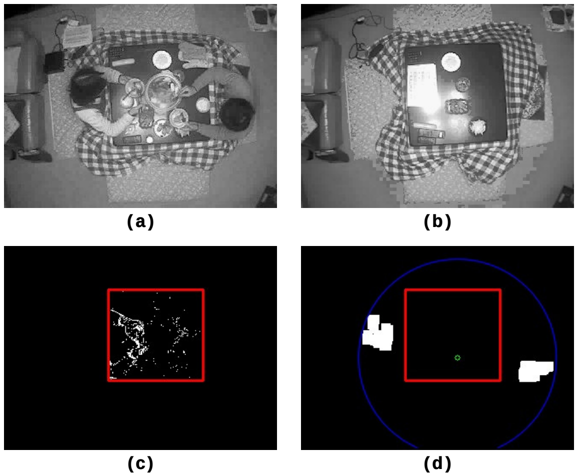

Figure 3. Visual tracking results. (a) shows typical input image. (b) shows updated background image. (c) shows tracked motion on the table. Red rectangle shows table area. White pixels shows changes due to motion. (d) White areas show occupant locations. Blue circle shows circle enclosing the occupants. Note that the table area was manually extracted.

Experiments

Method

We assessed the data collected to verify the effectiveness of our method. 50% of the 127 self-entered data points were randomly chosen and used as training data. The rest (50%) were used as test data. N (interval for calculating features) was set to 60 [seconds]. We investigated how well our method could predict the level of human interruptibility in the test data. *Pearson's product-moment correlation coefficient* was used to compared the values of the level of human interruptibility estimated by our method predicts and those given by the subjects.

Results
The results show that the correlation coefficient was 0.51. The main causes of incorrect predictions are as follows.

- One subject engaged in an action (such as sleeping) that differed from that of another occupant (who was writing).
- Different behaviors can yield the same physical attributes; for example, sleeping and watching TV yield very similar body motion.
- The same behavior may not be associated with the same level of interruptibility according to the time of day. Factors such as fatigue and physical condition may alter the level of interruptibility.

Discussion and Conclusions

This paper introduced a method for automatically estimating human interruptibility in a living room, based on audio-visual tracking. We conducted an extended experiment to verify the effectiveness of our method. The results show that the values of the level of human interruptibility our method estimates and those given by subjects are correlated to some extent. Although the correlation is not so high, we conclude that our method is practical for the following reason. Our method is simple to implement since it uses only one USB camera, a microphone, and one PC.

In the future, to improve the estimation accuracy, we will identify people in order to discriminate the intended partner from the others in the room, based visual sequential person tracking techniques. We will also investigate reasonable audio-visual sequential features other than static features to discriminate different behaviors that yield similar motions. We will consider contexts some as day of the week and time of day, and capture data over a much longer period such as one month and cover additional homes. Moreover, we will design an intuitive visual representation of interruptibility level such as color.

References

[1] Mcfarlane, D.C. Coordinating the Interruption of People in Human-Computer Interaction, *Proc. INTERACT 99'*, pp.295-303 (1999)

[2] Hudson, S. E., Fogarty J., Atkeson, C. G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J. C., and Yang, J. Predicting Human Interruptibility with sensors: A Wizard of Oz Feasibility Study, *Proc. CHI 03'*, pp.257-264 (2003)

[3] Minakuchi, M., Takeuchi, T., Kuramoto, I., Shibuya, Yu., and Tsujino Y. An Automatic Estimation Method for Busyness at Deskwork, *J. Human Interface Society* (In Japanese), Vol. 6, No. 1, pp. 69-74 (2004)

[4] Kidd, C. D., Orr, R., Abowd, G. D., Atkeson, C. G., Essa, I. A., MacIntyre, B., Mynatt, E., Starner, T. E. and Newstetter, W. The Aware Home: A Living Laboratory for Ubiquitous Computing Reseach, *COBUILD 99*', (1999)

[5] Yamazaki, T. Ubiquitous Home: Real-Life Tested for Home Context-Aware Service, *Proc. TRIDENTCOM 05'*, pp.54-59 (2005)

[6] Krumm, J., Harris, S., Barry, M., Brumitt, B., Hale, M., Shafer, S. Multi-Camera Multi-Person Tracking for EasyLiving, *VS 00'*, p.3 (2000)

[7] Burges, C. J. C. A Tutorial on Support Vector Machines for Patter Recognition, *Data Mining and Knowledge Discovery*, 2, pp.121-167 (1998)

[8] http://www.csie.ntu.edu.tw/~cjlin/libsvm/