PAPER

Improvement of Interruptibility Estimation during PC Work by Reflecting Conversation Status

Satoshi HASHIMOTO[†a)], **Takahiro TANAKA**[†], **Kazuaki AOKI**[†], ***Nonmembers***, ***and*** **Kinya FUJITA**[†], ***Member***

SUMMARY Frequently interrupting someone who is busy will decrease his or her productivity. To minimize this risk, a number of interruptibility estimation methods based on PC activity such as typing or mouse clicks have been developed. However, these estimation methods do not take account of the effect of conversations in relation to the interruptibility of office workers engaged in intellectual activities such as scientific research. This study proposes an interruptibility estimation method that takes account of the conversation status. Two conversation indices, "In conversation" and "End of conversation" were used in a method that we developed based on our analysis of 50 hours worth of recorded activity. Experiments, using the conversation status as judged by the Wizard-of-OZ method, demonstrated that the estimation accuracy can be improved by the two indices. Furthermore, an automatic conversation status recognition system was developed to replace the Wizard-of-OZ procedure. The results of using it for interruptibility estimation suggest the effectiveness of the automatically recognized conversation status.

key words: *interruptibility, desk work, conversation status, PC operation*

1. Introduction

Remote communication and online data access environments have improved with growth in computer usage and the Internet. In particular, improvements to remote information sharing environments should help to popularize distant working at satellite and home offices. On the other hand, the popularization of remote communication systems also means that there is a greater chance that workers will be interrupted by message alerts such as pop-up dialogue boxes on their PCs. However, most existing remote communication systems interrupt users without taking account of their work status, i.e., without determining whether they could be safely interrupted without disturbing their task. A prior study reported that frequent interruptions at inappropriate times can fragment thoughts of workers engaged in intellectual activities and decreases their productivity [1].

Various studies have attempted to solve this inappropriate interruption problem by developing a means to estimate the busyness of office workers. A group of studies based their development on the general tendency that the amount of keystrokes and mouse operations reflect the user's busyness when working on a PC (PC work) [2]. This approach is effective for work associated with physical activity. However, it is difficult to apply it to work not involving physical activity. On the other hand, interruptions at breakpoints during tasks where the worker switches from one subtask to another are expected to relieve cognitive workload, even when one is engaged in an intellectual activity [3]. Moreover, it has been shown that monitoring a combination of PC operations and head movements can improve the accuracy of interruptibility estimations for non-PC work as well as PC work [4].

Business conversations among workers occasionally occur in addition to the individual works. A previous study attempted to detect visitors and phone calls by using a camera and microphone [5]. It is considered that workers having conversations should not be interrupted so as not to disturb their communication [6]. Furthermore, a previous study on automatic estimation of interruptibility based on PC activity reported that 50 percent of the estimation errors were caused by conversations in the actual office work scenario [7]. This means the interruptibility estimation in the office environment needs to reflect the status of conversations that may be taking place as well as the individual work activity. The use of an omni-directional microphone in an office is a way of involving conversations in interruptibility estimation [8]. However, a wide variety of sounds, not just voices, are present in actual office environments, and a microphone would naturally pick them up as well. Therefore, what is needed is an appropriate way to discriminate conversations from other sounds and a way to incorporate it in the interruptibility estimation. We performed preliminary experiments using the Wizard-of-OZ (WoZ) method and reported the possibility of improving the estimation accuracy [9]. However, the effectiveness on different data sets and the feasibility of developing a system functioning in a real environment remained unclear.

In the current study, we studied the effectiveness of including a conversation status, as classified by the results of the WoZ method, in the interruptibility estimation by using two different data sets. Furthermore, we developed an automatic conversation status recognition system as a potential replacement for the Wizard-of-OZ procedure, and we experimentally investigated the estimation accuracy. The experimental results suggested that inclusion of the conversation status improves the estimation accuracy; even though the automatic conversation status recognition system has recognition errors.

Manuscript received March 27, 2014.
Manuscript revised August 19, 2014.
[†]The authors are with the Graduate School of Tokyo University of Agriculture and Technology, Koganei-shi, 184–8588 Japan.
a) E-mail: 50013646130@st.tuat.ac.jp
DOI: 10.1587/transinf.2014EDP7097

2. Interruptibility Estimation in Real Environments

2.1 Related Work

It is reported that interruptions fragment the thoughts of workers engaged in intellectual activities and decreases their productivity [1]. Therefore, it seems that an interruption control based on an estimate of the user's state would be useful for office workers who perform such intellectual work.

Various techniques have been investigated for estimating the user's state. One approach is to estimate the type of user activity, such as talking with a visitor or using a computer [5]. For these methods to work, sensors, such as microphones and cameras have to be installed in the workplace. The estimation relies on the relationship between interruptibility and type of activity. The problem with this method is that the temporal change in interruptibility is not principally reflected when the worker continues a single task. Thus, fine-grained control of the interruption timing appears difficult.

Another approach is based on PC operations such as keystrokes and mouse clicks [2], [10], [11]. These activities are useful metrics for physically observable tasks. However, physical activities do not necessarily reflect intellectual activities, which should not be interrupted. Therefore, the effect of breakpoints in tasks, where one changes from one subtask to another, has been studied, and it has been demonstrated that interruptibility instantaneously increases at such breakpoints even when the user is engaged in an intellectual activity [3]. Furthermore, an interruptibility estimation method has been proposed on the basis of focused application-switching (AS), which is a potential alternative to a breakpoint estimation in a real environment [7], [12].

The above-mentioned interruptibility estimation methods commonly depend on task-related activities. However, conversations between workers occasionally occur in offices, and interruptibility during a conversation tends to be low [6] because people prefer their conversations not to be disturbed. The problem is that in order to engage in conversations in an office, workers would naturally pause their operations on their PCs. Therefore, the decrease in PC operation activity leads the interruptibility of this situation to be erroneously estimated as high [7].

2.2 Effect of Conversation Status on Interruptibility

As described above, an estimation method based only on PC operations mistakenly estimates conversation states as highly interruptible. In fact, 50 percent of the estimation errors can be attributed to conversations [7]. Therefore, the detection and proper reflection of conversation is expected to improve the estimation accuracy.

On the other hand, office conversations, such as discussions and consultations, are part of most jobs, as are PC operations. Hence, the end of a conversation can be regarded as a switch from a task to another task. Therefore, the interruptibility of the worker is speculated to temporarily increase from the end of the conversation to the beginning of the next task, which is similar to the case of AS during PC operation. However, this assumption needs to be verified. In addition, the importance of the topic of conversation, such as its relevance to the business at hand, is also inferred to affect interruptibility. However, the automatic estimation of the purpose, content and importance of conversation is as yet a challenge.

We experimentally investigated the relation between interruptibility and conversation statuses classified by using the WoZ method. Furthermore, we examined the feasibility of improving the interruptibility estimation accuracy by incorporating two conversation statuses in the estimation method based on PC operation.

3. Interruptibility Estimation Method Reflecting Conversation Status

This section analyzes the relation between interruptibility and conversation status on the basis of the results of interruption experiments conducted on participants while they were working. After that, we develop an extended interruptibility estimation method that accounts for the conversation status as well as PC operations.

Similar to the previous studies [2], [3], [7], [10], we target office workers, who perform intellectual activities at their own desk on PCs. Such work could include, for example, R&D and certain types of clerical work. Furthermore, we assume an environment where two to more than a dozen people work in an office and occasionally engage in conversations.

3.1 Interruption Experiments

We conducted a set of experiments, in which participants were interrupted while they were working. The experimental system is shown in Fig. 1 (a). In the experiments, we recorded the PC operation information and the sound. In addition, the system interrupted the participant automatically by using a sound and a dialog box when one of the pre-programmed interruption conditions was satisfied. The participants were requested to score their subjective interruptibility in the dialog box.

The recorded operation information was keystrokes, mouse clicks, mouse wheel movement, active window name, process ID, window message (quit, clipboard), and number of opened windows. The sample interval was set to 500 ms. We installed an omni-directional microphone on the PC monitor, as shown in Fig. 1 (b) and recorded the sound at 22050 samples per second. The participants were requested to score their own interruptibility on a five-grade scale, from 1 (absolutely uninterruptible) to 5 (absolutely interruptible).

The following simple interruption rules were set to allow the experimental system to automatically collect interruptibility scores in "potential conversation" and "potential

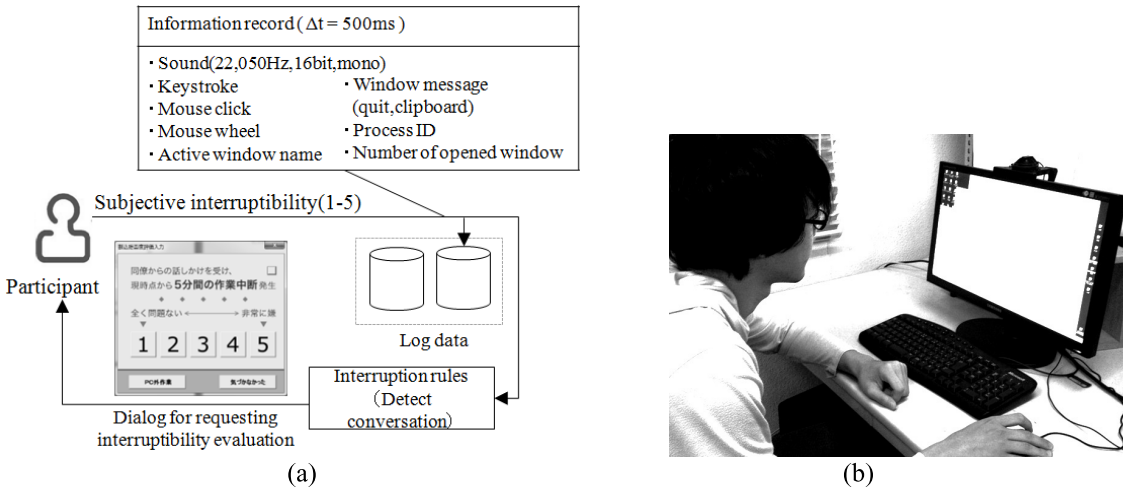


Fig. 1 Experimental system. (a) configuration (b) example of experimental scene.

conversation ending" states.

1. Continuous sound detection (potential conversation): The rate of samples in which the sound pressure exceeded the threshold is between 25 to 90 percent in the last 20 seconds, and more than 15 percent in the first half of the period.
2. End of continuous sound (potential conversation ending): Condition 1 was satisfied in the previous sample but not in the current sample.
3. Quiet period (potentially not a conversation): Other than conditions 1 and 2.

The threshold for each condition was determined on the basis of preliminary experiments. The threshold of sound pressure was set to the central value of the averages of 4-second environmental sound and eight peaks in 4-second conversational voices prior to the experiment. The period for the continuous sound detection was set to 20 seconds in order to reduce the influence of short utterances such as greetings and silent durations in conversational turn-taking. In real office environments, very short conversations, such as greetings, also occur. However, it is speculated that their influence on interruptibility is smaller than continuous conversations. Therefore, detection of short conversations was avoided by using the detection rate in both the first half and the entire part of the 20-second period. Furthermore, minimum interruption intervals, 300 seconds for continuous sound and 900 seconds for quiet periods, were imposed on each condition in order to avoid excessive numbers of interruptions and their mental bias on the subjective scores. The minimum interruption interval for the end of a conversation was set to 30 seconds because the chance to interrupt at the end of the conversation is limited.

As already mentioned, the environment is assumed to be one where two to more than a dozen people work on PCs at their own desk and engage in occasional conversations. In keeping with this assumption, the experiments were performed at university laboratories having four to ten people. These labs had no machinery that generated loud or continuous noise.

The participants were two faculty members and eight university students in the department of computer and information sciences. The people in the room other than participants were also faculty members and university students. All the participants and room members engaged in research on their PCs. We did not impose any constraints on their conversations, work content, or other activities. We instructed the participants to perform their activities as usual, with exceptions being that we requested them not to go out for a long time and not to play music or video.

Each participant engaged in a five-hour experiment at their own desk. The main observed activities were reading and writing documents such as journal papers and reading and writing e-mails. Conversations were on their research, news, and hobbies.

3.2 Experimental Results

In reviewing the recorded data, the automatically classified conversation statuses using simple rules were different from the actual states of conversations in several cases. Therefore, we re-classified the experimental data using the WoZ method based on the following rules.

1. In conversation: a human voice has been detected in the last 10 seconds.
2. End of conversation: a human voice was detected in the last 20 seconds, but not in the period from 10 seconds before to 10 seconds after the interruption. Cases in which keystrokes or mouse clicks were detected were excluded because the PC operation implies the participant resumed his or her task.
3. Quiet period: Other than conditions 1 and 2.

The average interruptibilities of the three conversation statuses are shown in Fig. 2. The numbers of interruptions "in conversation", at the "End of the onversation", and during "Quiet periods" were 310, 78, and 226, respectively. The

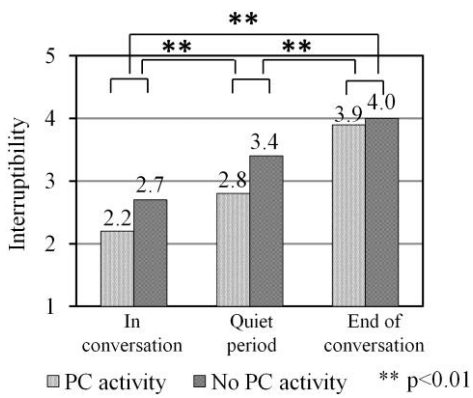


Fig. 2 Average interruptibility for each conversation status.

average interruptibility was 2.8.

The previous studies pointed out that PC operation affects interruptibility. Hence, the data was categorized according to the PC activity in addition to the conversation status. The existence of PC activity was judged on the basis of detected keystrokes or mouse clicks within the last 30 seconds.

The results of a two-way analysis of variance (ANOVA) between conversation status and PC operation revealed that they significantly affect interruptibility [conversation status: $F(2,609)=40.$ $p<0.01$], [PC operation: $F(1,609)=8.3.$ $p<0.01$]. As a result of multiple-comparisons with Bonferroni's correction, interruptibility during a conversation was significantly lower and interruptibility at the end of a conversation was significantly higher than during quiet periods ["In conversation" and "Quiet period": $p<0.01$], ["In conversation" and "End of conversation": $p<0.01$], ["Quiet period" and "End of conversation": $p<0.01$]. These results show that the "In conversation" and "End of conversation" statuses are promising as indices for the interruptibility estimation. Furthermore, the significantly lower interruptibility for the data with PC operation suggests the influence of conversations varies in accordance with PC operations. It appears that the reflection of the conversation status on the interruptibility estimation should take account of the influence of PC operations as well.

3.3 Extension of Interruptibility Estimation Algorithm

On the basis of the results presented in 3.2, we devised an interruptibility estimation algorithm by adding conversation indices to the estimation method based on PC operation. The baseline method estimates interruptibility using two individual algorithms for the moments at the switching of a focused application software (Application-Switching: AS) and for the periods while one application software is continuously used (Not Application-Switching: NAS) [7]. During NAS period, the estimation method uses four indices relevant to PC operation activity [2], [10]. At AS moments, interruptibility is affected by the relevance of the tasks before and after AS, in addition to the operation activity. In total, the estimation method utilizes 19 indices related to task relevance and PC operation activity [12].

Table 1 Indices of the interruptibility estimation during NAS.

ID	Indices	Interruptibility
A	Keystroke in last 20 s	Low
B	PC activity detected in more than 30 percent of the last 2 min.	Low
C	Use of both keyboard and mouse in the last 2 min.	Low
D	Transitioned from shell (desktop) within 5 min.	Low

Table 2 Indices based on conversation status.

ID	Indices	Interruptibility
E	In conversation	Low
F	End of conversation	High

However, AS rarely occurs during conversations, because users usually stop operating their PCs while in conversation. Therefore, we introduced two conversation indices into the estimation equation for NAS. Equation (1) represents the estimation equation for NAS. The equation is composed of four indices, as shown in Table 1. The index takes the value 1 when the condition for each index is satisfied, and 0 otherwise. The value of function f ranges from 0 to 1. The interruptibility is classified into three levels using Eq. (2).

$$f = (2\mathrm{A} + \mathrm{B} + \mathrm{C} + \mathrm{D})/5 \tag{1}$$

$$Interruptibility = \begin{cases} Low & 0.7 \le f \le 1 \\ Medium & 0.2 \le f < 0.7 \\ High & 0 \le f < 0.2 \end{cases} \tag{2}$$

The added conversation indices are shown in Table 2. In 3.2, the results of an ANOVA suggested that the effect of "In conversation" and "End of conversation" on interruptibility together with the existence or absence of PC activity. Therefore, we defined two estimation equations for each case (Eq. (3) and (4)). After the calculation of function f, the interruptibility was estimated in three levels using Eq. (2).

PC activity :

$$f_{PC} = (2\mathrm{A} + \mathrm{B} + \mathrm{C} + \mathrm{D} + \mathrm{E} + 2\overline{\mathrm{F}})/8 \tag{3}$$

No PC activity :

$$f_{NPC} = (2\mathrm{A} + \mathrm{B} + \mathrm{C} + \mathrm{D} + 2\mathrm{E} + \overline{\mathrm{F}})/8 \tag{4}$$

In comparison with the original estimation Eq. (1), the effects of the PC operation indices in Eq. (3) and (4) are relatively smaller, since the equations have more indices. Therefore, the same threshold has different effects on Eq. (3) and (4) from those on Eq. (1). On the other hand, the purpose of this study is to investigate the effect of the conversation status itself on interruptibility, not the thresholds. Therefore, we set the thresholds and the coefficients for indices A-D in

Eq. (3) and (4) to the same values as in Eq. (1). A further discussion on the thresholds and coefficients is in 6.2. The coefficients for E and F were determined by manually exploring integer values to see which ones provided the highest precision for the analysis data set.

4. Automatic Conversation Status Recognition in Real Environment

A means of automatic recognition is needed in order to incorporate the conversation status into an interruptibility estimation system. However, even with such a system, the estimation accuracy might not improve since recognition errors are unavoidable. Therefore, we developed an automatic conversation status recognition system and experimentally examined its effect on the estimation accuracy.

4.1 Automatic Conversational Voice Detection

Conversation recognition requires a two-stage algorithm. The first stage is the conversational voice detection. The second stage is the recognition of whether the detected voice is a part of conversational utterance or not. Numerous automatic human voice detection algorithms have been proposed, including ones based on periodicity [13], power ratio in the frequency domain [14], and frequency deviation [15]–[17].

To correctly recognize the conversation period and the end of the conversation, the voice of the talking partner, not only the target person of the estimation, has to be detected. In this case, as the distance from the microphone increases, the voice detection function gets more susceptible to background noise. In addition, in real office environments, noises that have periodicity and frequency components close to those of human voices exist; these noises include the ringers of phones and fricative sounds of objects on motion. Therefore, we attempted to detect voice activity using the following method based on the features of the human voice and by referring to previous studies.

Figure 3 shows an example of a sound pressure waveform of a conversational voice lasting 2.5 seconds together with its spectrogram obtained by wavelet transform. Conversational voices have pitch frequencies ranging from 80 to 180 Hz for men and from 160 to 280 Hz for women [18]. The pitch and its harmonics form a striped pattern [19]. Moreover, the pitch frequency deviation within a mora is smaller compared with the frequency change between mora [20]. Therefore, we used harmonicity, the pitch deviation rate, and the sound pressure conditions for the human voice detection.

The voice detection procedure is shown in Fig. 4. The detection is performed at 0.5 second intervals. Input sound is divided into 160-ms bins and transformed with a continuous wavelet transform at 1/128th second intervals. The pitch and the second harmonic frequencies are detected by detecting the peaks in the frequency domain. The pitch frequency is searched within a range of 80 to 360 Hz. After the calculation of the average for 0.5 seconds and the variances for eight subsections, the sample is detected as a conversational voice if the following four conditions are satisfied.

1. Harmonicity condition 1 (the existence of second harmonic): A peak exists in the range from 1.68 to 2.37 times the pitch frequency.
2. Harmonicity condition 2 (clarity of stripe pattern): The powers of both the pitch and the second harmonic are more than ten times larger than the powers of the intermediate frequency of the pitch and the second harmonic.
3. Pitch deviation condition: The number of subsections whose pitch variance is less than 40 Hz is more than 3.
4. Sound pressure condition: The sound pressure is at least 1.2 times that of the environmental sound.

The thresholds for each condition were experimentally determined and by referring to previous studies.

4.2 Automatic Conversation Status Recognition

The automatic conversation status recognition algorithm uses a voice detection rate similar to the interruption rules for the experiment, in consideration of short conversations

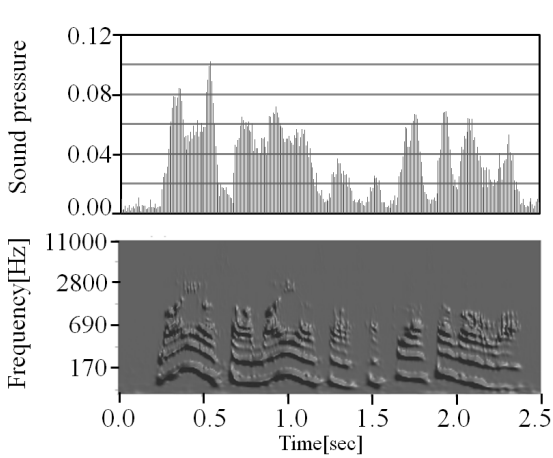


Fig. 3 Sound pressure and spectrogram of conversational voice.

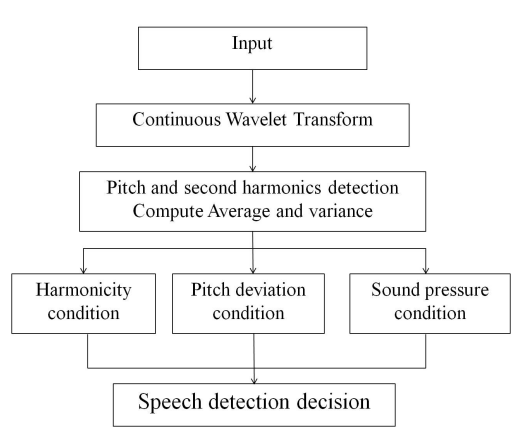


Fig. 4 Conversational voice detection algorithm.

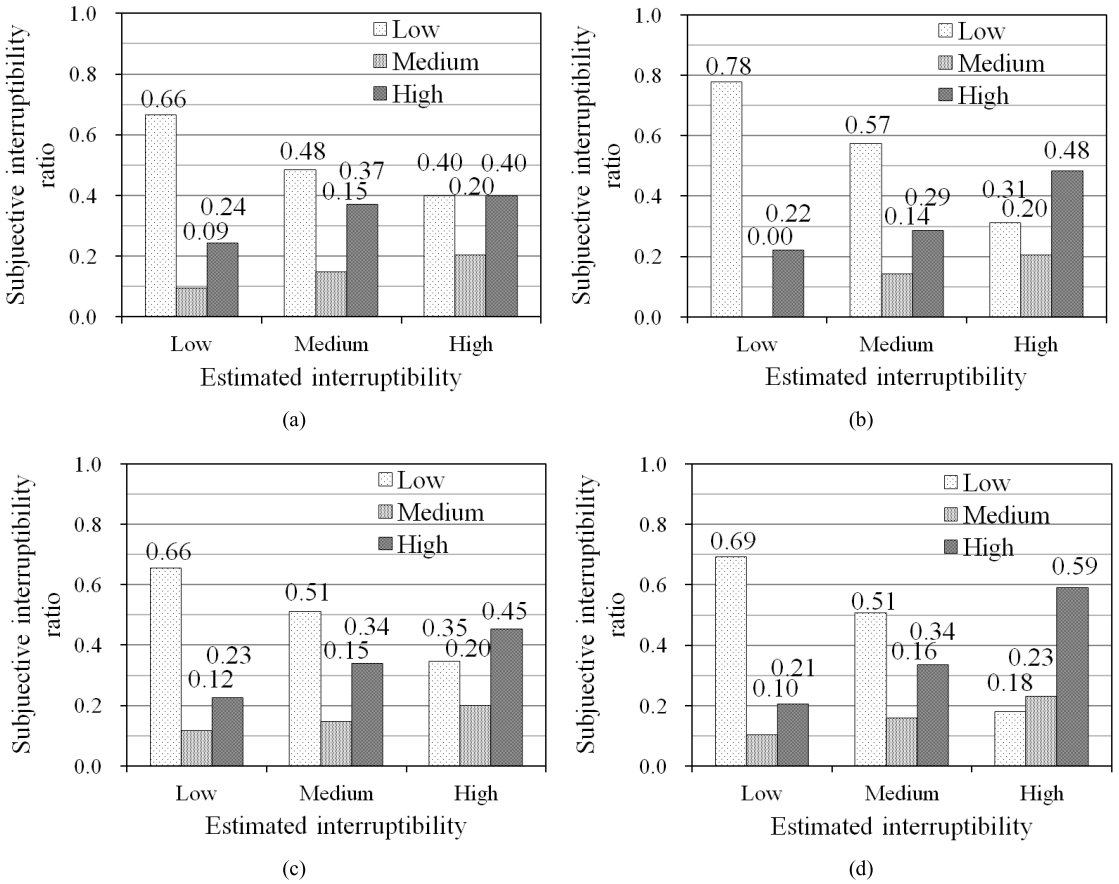


Fig. 5 Interruptibility estimation results: (a) previous method (b) with "In conversation" (c) with "End of conversation" (d) with both indices.

such as greetings and silent durations in conversational turn-taking. The threshold of the voice detection rate for the "In conversation" state was lowered from 25 percent to 15 percent to compensate for voice detection failures.

1. In conversation: The rate of samples whose sound pressure exceeded the threshold is between 15 to 90 percent in the last 20 seconds, and more than 15 percent occur in the first half of the period.
2. End of conversation: Condition 1 was satisfied in the previous sample but not in the current sample.
3. Quiet period: Other than conditions 1 and 2.

5. Results of Interruptibility Estimation

We examined the possibility of using the conversation status to improve the accuracy of the interruptibility estimation.

5.1 Estimation Results Using WoZ Method

We applied the proposed interruptibility estimation method to the analysis data set that was used for the estimation indices selection. The conversation status of the data was classified using the WoZ method. The accuracy was evaluated on the basis of the ratio of the subjective interruptibility scores. For comparison, the five-level subjective evaluation scores that participants evaluated were converted into three-level scores. Figure 5 (a) shows the results of the estimation using the previously proposed PC operation-based indices. Figure 5 (b, c, d) shows the offline estimates for the inclusion of the "In conversation", "End of conversation", and both indices, in addition to the PC operation-based indices.

As shown in Fig. 5 (a), the precision of the previous method for high interruptibility was 0.40. It appears that decreased PC operation activity during a conversation caused the interruptibility to be evaluated as high. The F value that is the average of precision and recall was 0.38. On the other hand, inclusion of the "In conversation" index improved the precision for high interruptibility to 0.48. In particular, the high-risk estimation error, which is the error of estimating a low interruptibility as high, decreased to 0.31, and the F value increased to 0.41.

As speculated in 2.2, the inclusion of the "End of conversation" state also improved the precision for high interruptibility to 0.45 and the F value to 0.40, respectively. The application of both indices had a greater effect than the application of each individual index; using both improved the precision to 0.59. The high-risk estimation error rate decreased to 0.18, which implies a reduction in the work inhibition risk. The PC operation-based method principally

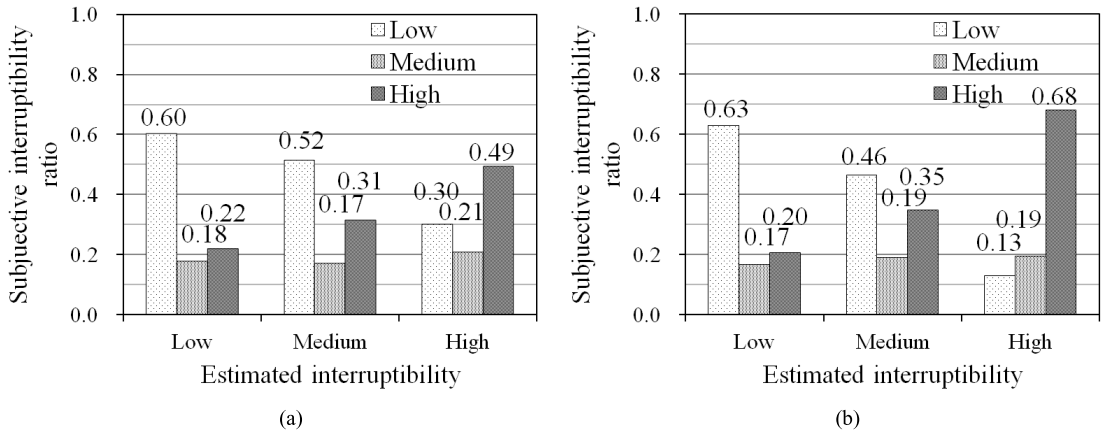


Fig. 6 Interruptibility estimation results using evaluation data set: (a) previous method and (b) with conversation indices.

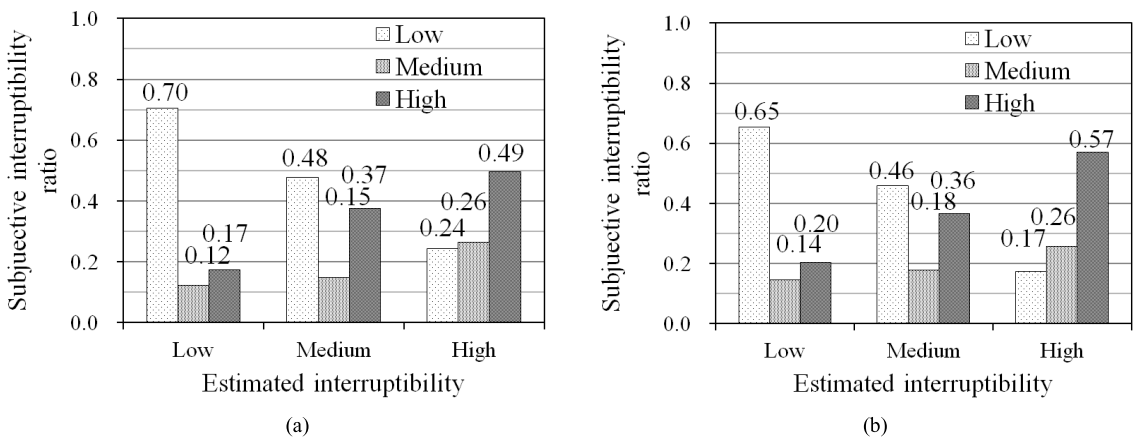


Fig. 7 Interruptibility estimation results using automatic conversation recognition: (a) analysis data set and (b) evaluation data set.

views a decrease in PC operation activity caused by conversation as work stagnation. It appears this sort of error was reduced by including the conversation status. In addition, the precision for low interruptibility was higher than with the PC operation-only method. These results imply that the conversation status improves the estimation accuracy especially for high interruptibility.

5.2 Estimation Results for Different Data Set

The interruptibility estimation accuracy was examined for a different data set that was collected in another set of experiments. The experimental system and the conditions were the same as in the previous experiments. We recorded 5-hour data, each of two faculty members and eight students. The conversation status was classified by WoZ method again.

Figures 6 (a) and (b) are respectively the estimation results using the PC operation-based indices only and the PC-based and conversation indices. Similar to the results for analysis data set, the precision for high interruptibility improved from 0.49 to 0.68 as a result of including the conversation status. The high-risk estimation error decreased from 0.30 to 0.13. The F value also improved from 0.40 to 0.45. These accuracy improvements suggest the effectiveness of the two conversation indices on the interruptibility estimation.

5.3 Estimation Results Using Automatic Conversation Status Recognition

By applying WoZ method, the effect of including the conversation status on estimation accuracy was demonstrated. In this section, we discuss the feasibility of the estimation accuracy improvement by applying the automatic conversation status recognition method described in the previous section.

Figure 7 shows the interruptibility estimation results based on the automatic conversation status recognition. Figure 7 (a) and (b) represent the results for the analysis data set and the evaluation data set. In comparison with the PC operation-based method shown in Fig. 5 (a) and Fig. 6 (a), the precision for high interruptibility are improved in a similar way to the results for the WoZ method. The rate of high-risk estimation errors decreased by more than 0.10 in both

Table 3 Results of conversation recognition (number of samples).

		Wizard-of-OZ		Precision
		conver-sation	No conv-ersation	
Automatic recognition	Conver-sation	243	70	0.78
	No Conv-ersation	67	222	0.77
Recall		0.78	0.76	

data sets. The F value also improved to 0.40 in the analyzed data and 0.42 in the evaluation data set. On the other hand, in comparison with the results for the WoZ method in Fig. 5 and Fig. 6, the estimation accuracy was lower; the precision for high interruptibility was 0.49 in the analysis data set, and 0.57 in the evaluation data set. The improvement effect was about the half of the interruptibility estimation results using the WoZ method. It appears that the lower precisions are due to errors in recognizing the conversation status.

Table 3 shows the results of the automatic conversation status recognition for the analysis data set. The average precision and recall of the conversation status recognition were 0.77 and 0.77. An improvement in conversation status recognition accuracy may be able to improve the interruptibility estimation accuracy to a level equivalent to that of the WoZ method. The causes and the feasibility of such an improvement will be discussed in 6.1. Although the automatic conversation status recognition algorithm still has estimation errors, the improvement of the estimation accuracy was demonstrated. It implies that automatic recognition and inclusion of a conversation status is a promising way to improve the accuracy of the interruptibility estimation.

6. Discussion

6.1 Feasibility of Improving the Accuracy of Automatic Conversation Status Recognition

In the interruptibility estimation based on the automatic conversation status recognition, the precision for high interruptibility was improved and the rate of high-risk errors was decreased. However, in comparison with the results using the WoZ method, the improvement effect was about half. As shown in Table 3, the accuracy of the automatic conversation status recognition was less than 80 percent. Since the two conversation indices have opposite effects on interruptibility as revealed in the experiments, it appears the lower estimation accuracy was due to the errors in detection between the two indices.

Therefore, we manually checked the incorrectly recognized sections. The major cause was the failure to detect the voice of the conversational partner when environmental noise such as from an air conditioner decreased the S/N ratio. In addition, failures to detect the pitch and harmonics occurred when more than one person spoke simultaneously. The deterioration due to the former cause can be alleviated by performing the voice detection after the noise component in the signal is suppressed [21]. The latter cause might be improved by detecting individual voice features for each speaker.

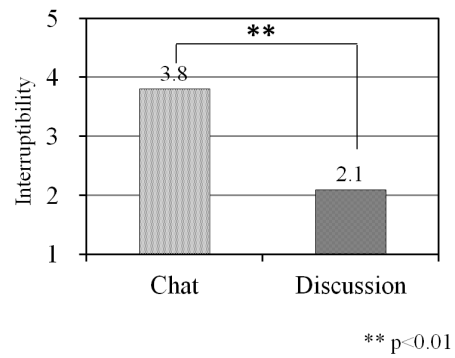


Fig. 8 Relationship between conversational content and interruptibility.

6.2 Feasibility of Interruptibility Estimation Accuracy Improvement

The use of two states, i.e., "In conversation" and "End of conversation", improved the accuracy of the interruptibility estimations for office work with occasional conversations. The results of the experiment also support the hypothesis that interruptibility temporarily increases at the end of a conversation.

In this study, to examine the effect of the conversation status alone, the thresholds and the coefficients for indices A-D were set to the same values as in the previous method described in 3.3. However, the used thresholds and coefficients were obviously not optimal because the number of indices is large and the proposed algorithm distinguishes situations depending on the existence of PC activity. Accordingly, we set the threshold at 0.125 for high interruptibility. Setting a high interruptibility threshold improved the precision for both data sets. Despite this, the recall and F value decreased for both sets. If we are to make further improvements to the estimation accuracy, the thresholds and coefficients need to be examined.

It was confirmed that interruptibility during a conversation tends to be low. However, in the experiments of this study, some participants said their interruptibility was high during some conversations. They reported that they were at ease being interrupted when the conversation was not important. This suggests that interruptibility varies depending on the importance of the conversation.

Since an interruptibility variation depending on conversational content was suggested, we manually classified the conversational content at the moments of the interruptibility evaluation in the analysis data set into "casual chat" or "serious discussion", and compared the interruptibility scores for the two categories, as shown in Fig. 8. The number of interruptions during casual chats was 60 out of 310 interruptions. The average interruptibility was 3.8. On the other hand, the number of interruptions during serious discussions was 250, and their average interruptibility was 2.1. A sig-

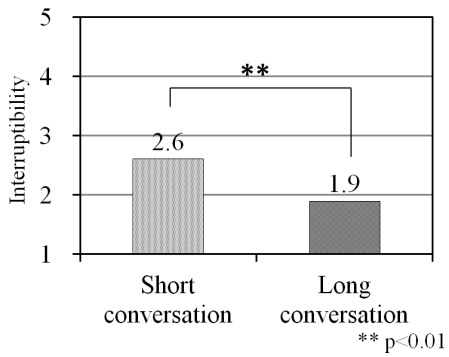


Fig. 9 Relationship between conversation duration and interruptibility.

nificant difference was observed between them in Welch's t-test [t(101)=2.0, p<0.01]. This suggests that interruptibility differs depending on the conversational content. Therefore, the estimation accuracy is expected to be improved by developing an automatic recognition technology of the conversational content.

An effective recognition method of the conversational content will require speech recognition and text processing technologies. However, a substitutive approach such as detection and counting of instances of laughter based on phonological features might be feasible. As an alternative, we can consider the conversation duration. A low interruptibility conversation, such as an important conversation or serious discussion, is expected to be longer. Accordingly, we analyzed the correlation between the interruptibility and the elapsed time from the beginning of the conversation. The results showed weak negative correlation. Therefore, we defined two conversation types, "Short conversation" and "Long conversation". The threshold for dividing the two types was set to 120 seconds based on the results of a cluster analysis. The interruptibility scores for the two conversation types are shown in Fig. 9. The average interruptibilities of "Short conversations" and "Long conversations" were 2.6 and 1.9, respectively. A significant difference between them was confirmed by Welch's t-test [t(103)=2.0, p<0.01]. These results suggest the effectiveness of the conversation duration on improving the accuracy of the interruptibility estimation. Further improvements can be expected by devising a method of estimating the importance of a conversation.

7. Conclusions

This study is a step toward the goal of realizing an interruptibility estimation that will work in actual office environments. Here, we focused on the conversation status, especially the possibility of a temporary increase in interruptibility at the end of a conversation. The experiments based on the WoZ method demonstrated the improvement in accuracy had by incorporating this conversation status into the interruptibility estimation. After that, we developed an automatic conversation status recognition system based on the voice detection rate and confirmed that it improved the estimation accuracy on actual data sets. Future work will include improving the recognition accuracy of the conversation status, an exploration of the optimal thresholds and coefficients, and a study on how to reflect the conversational content and importance in the interruptibility estimation.

Acknowledgements

This work was partly supported by a grant from the Japan Society for the Promotion of Science (JSPS), funds for "smart space technology toward sustainable society" from MEXT, and funds for "ultra-realistic communication systems" from NICT.

References

[1] B.P. Bailey, J.A. Konstan, and J.V. Carlis, "The effects of interruptions on task performance, annoyance, and anxiety in the user interface," Proc. INTERACT'01, vol.1, pp.593–601, July 2001.
[2] M. Minakuchi, T. Takeuchi, I. Kuramoto, Y. Shibuya, and T. Tsujino, "An automatic estimation method for busyness at deskwork," Trans. Hum. Interf. Soc., vol.6, no.1, pp.69–74, Feb. 2004. (in Japanese).
[3] S.T. Iqbal and B.P. Bailey, "Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption," Proc. ACM CHI'05, pp.1489–1492, April 2005.
[4] R. Abe, T. Tanaka, and K. Fujita, "Study of uninterruptibility estimation based on head motion and PC operation records during desk work," Corres. Hum. Interf. Soc., vol.13, no.10, pp.15–22, Nov. 2011. (in Japanese)
[5] J. Fogarty, S.E. Hudson, C.G. Atkeson, D. Avrahami, J. Forlizzi, S. Kiesler, J.C. Lee, and J. Yang, "Predicting human interruptibility with sensors," ACM Trans. CHI, vol.12, no.1, pp.119–146, March 2005.
[6] D. Siewiorek, A. Smailagic, J. Furukawa, N. Moraveji, K. Reiger, and J. Shaffe, "SenSay: A context-aware mobile phone," Proc. IEEE ISWC'03, p.248, Oct. 2003.
[7] T. Tanaka, S. Fukasawa, T. Kanamaru, M. Nonaka, and K. Fujita, "Study of uninterruptibility estimation method for office worker during PC work," Proc. Hum. Interf. Symp.'11, pp.333–338, Sept. 2011. (in Japanese)
[8] K. Aoki, S. Sato, T. Tanaka, and K. Fujita, "A feasibility of office interruptibility estimation based on ceiling camera images and sounds," Proc. IEEE VR'13, pp.1–4, March 2013.
[9] S. Hashimoto, T. Tanaka, K. Aoki, and K. Fujita, "Estimation of interruptibility during office work based on PC activity and conversation," Proc. HCI International 2013, vol.3, pp.297–306, July 2013.
[10] S. Honda, H. Tomioka, T. Kimura, T. Oosawa, K. Okada, and Y. Matsushita, "A home office environment based on the concentration degrees of workers — A virtual office system "Valentine"," IPSJ J., vol.39, no.5, pp.1472–1483, May 1998. (in Japanese)
[11] K. Shimizu, K. Yamashita, K. Nishimoto, and S. Kunifuji, "A system that provides personal workload awareness using character agent," Proc. Jpn Soc. Artif. Intell., 1B3-03, pp.1–3, June 2004. (in Japanese)
[12] T. Tanaka and K. Fujita, "Study of user interruptibility estimation based on focused application switching," Proc. CSCW'11, pp.721–724, March 2011.
[13] K. Ishizuka and T. Nakatani, "Study of noise robust voice activity detection based on periodic component to aperiodic component ratio," Proc. ISCA SAPA'06, pp.65–70, Sept. 2006.
[14] C.F. Juang, C.N. Cheng, and T.M. Chen, "Speech detection in noisy environments by wavelet energy-based recurrent neural fuzzy network," Expert Sys. App., vol.36, no.1, pp.321–332, Jan. 2009.
[15] H. Lu., W. Pan, N.D. Lane, T. Choudhury, and A.T. Campbell,

"SoundSense: Scalable sound sensing for people-centric applications on mobile phones," Proc. MobiSys'09, pp.165–178, June 2009.

[16] L. Lu, H. Jiang, and H.J. Zhang, "A robust audio classification and segmentation method," Proc. ACM Multimedia'01, pp.203–211, Oct. 2001.

[17] J.L. Shen, J.W. Hung, and L.S. Lee, "Robust entropy-based endpoint detection for speech recognition in noisy environments," Proc. ICSLP'98, pp.232–235, Dec. 1998.

[18] Y.M. Zeng, Z.Y. Wu, T. Falk, and W.Y. Chan, "Robust GMM based gender classification using pitch and RASTA-PLP parameters of speech," Proc. ICMLC'06, pp.3376–3379, Aug. 2006.

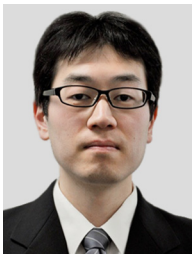
[19] Y. Nomura, J. Lu, H. Sekiya, and T. Yahagi, "Speech enhancement based on the dominant classification between speech and noise using feature data in spectrogram of observation signal," IEEJ Trans. EIS, vol.124, no.11, pp.2310–2319, Nov. 2004. (in Japanese)

[20] K. Iwano and K. Hirose, "A statistical modeling of fundamental frequency contours in moraic unit and its use for the detection of prosodic word boundaries," IPSJ J., vol.40, no.4, pp.1356–1364, April 1999. (in Japanese)

[21] K. Ishizuka, M. Fujimoto, and T. Nakatani, "Advances in voice activity detection," Acoust. Sci. Tech., no.65, vol.10, pp.537–543, Oct. 2009. (in Japanese)

Satoshi Hashimoto received his B.S. degree in the Department of Computer and Information Sciences from Tokyo University of Agriculture and Technology, Japan, in 2013. He is currently pursuing a Master's course at the university. His research interest is user state estimation using conversational voice information.

Takahiro Tanaka received his Ph.D. degree in Engineering from Tokyo Institute of Technology, Japan in 2006. He is currently an Assistant Professor in the Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology. His research interests include intelligent agents and human-computer interaction.



Kazuaki Aoki received his Ph.D. in Computer Science from Hokkaido University, Japan in 2011. He is currently an Assistant Professor in the Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology. His research interests include pattern recognition, machine learning, and human interfaces.

Kinya Fujita received his Ph.D. degree from Keio University, Japan, in 1988. After working at the Sagami Institute of Technology, Tohoku University, and Iwate University, he joined Tokyo University of Agriculture and Technology. He is currently working as a Professor in the Department of Computer and Information Sciences. His research interest is the development of smart human interface systems including interruption controls, remote communication systems, and haptics in virtual reality. He is a member of IEEE, ACM, IPSJ, and SICE, and a board member of HIS and VRSJ.