Analysis of Customer Communication by Employee in Restaurant and Lead Time Estimation

Masanori Takehara*, Hiroya Nojiri*, Satoshi Tamura*, Satoru Hayamizu* and Takeshi Kurata†
*Department of Engineering, Gifu University, Gifu, Japan
E-mail: (takehara, nojiri)@asr.info.gifu-u.ac.jp, tamura@info.gifu-u.ac.jp, hayamizu@gifu-u.ac.jp
†Center for Service Research, National Institute of Advanced Industrial Science and Technology, Ibaraki, Japan
E-mail: t.kurata@aist.go.jp

***Abstract*—Human behavior sensing and their analysis are great role to improve service quality and education of employees. This paper shows novel frameworks of detection of customer communication and lead time estimation(LTE) by using multi-sensored data, sound data and accounting data in the restaurant. They are useful for management about work environments and problems for employees. Lead time from order to delivery shows the quality of the service for customers. We found sound data of an employee's speech is useful for these techniques by speech ratio smoothing and POS sound detection.**

I. Introduction

Recently, human behavior estimation of employees becomes one of the attractive themes related with signal processing and pattern recognition fields. In order to visualize employees' operations, sensor and image data of employees have been acquired in some workspaces such as factory[1] and hospital[2].

The goal of this paper is to estimate lead time for customers in a restaurant. The lead time means the time from ordering a meal to serving. In other words, it is waiting time for their meals. The lead time is one of useful indicators for a restaurant because it shows the quality of the service. If lead times are getting shorter, it leads to enhancement of customer satisfaction. The lead time is often predicted by a number of customers or orders. We propose a novel approach to estimate the lead time automatically. If we can know when an employee takes order and serves meals for customers, we can estimate the lead time by taking a difference between these actions. So we focus on an employee's operations and detect them using sound data and locational data of the employee. If we can know the lead time more correctly, the lead time is appropriate to grasp a problem of services and improve the employee's work.

This paper is organized as follows: Related and previous works are described in Section 2. Waiting operation detection method and experiment are explained in Section 3. An experiment of lead time estimation(LTE) and result are shown in Section 4. Finally we conclude this paper in Section 5.

II. Previous Works

In our researches, we have conducted 3 steps in order to support to improve employees' works, work environment and education for them. First of all, we measure behavior of employees such as location, motion and speech. Also we collect accounting data of customers. Next, we perform basic analysis to these data. Then, we apply the result to management analysis such as LTE and visualize.

A. Data Collection

We collected behavior data of employees in a restaurant by using PDR(Pedestrian Dead-Reckoning) sensors and bone-conductive microphones. Additionally, we gathered accounting data of customers from POS(Point-of-Sales) terminals. The detail of data is shown in [3]. A PDR sensor enables us to record triaxial acceleration and orientation so that we can estimate the position coordinate of the employee. A bone-conductive microphone enable us to record speech of the employee.

We measured these data for about three months. We aim to observe difference of their working behavior by long term measurement. The total amount of recorded data is roughly 3,000 hours. Sixteen employees took part in the measurement.

B. Basic Analysis of Sound Data

We have already analysed sound data of employees. Voice activity detection(VAD) is one of basic sound analysis techniques. We can know when the employee who equipped a microphone spoke and how long the employee talked. Speech ratio means the rate how often the employee speak and keep silent. It is utilized to service operation estimation[4].

In this paper, we chose low-frequency feature and harmonic structure-based features[5] for VAD method because of two reasons. First, in bone-conductive sound, low-frequency power is concentrate while the employee who equipped a microphone is speaking compared with air-conductive sound. Second, the noise when a user touches his/her microphone often caused in bone-conductive sound. Harmonic structure-based features are effective in order to prevent these noises.

Utterance classification is a method applied with speaker identification technique. In a restaurant, location of a speaker who equipped a microphone determines which the speaker talks with other employees or customers. The paper [6] showed the accuracy of the utterance classification can be improved by using not only sound data but also locational data simultaneously.

C. Lead Time Estimation

The results of basic analysis are difficult to use for employees and their managers in a restaurant. They want indicators

978-616-361-823-8 © 2014 APSIPA

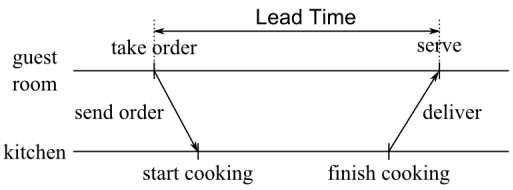

Fig. 1. The lead time and operations of employees.

that enable to evaluate their works or the qualities of the services. We should focus on not only the results but also what the results mean in their services. For example, staying ratio and speech ratio of employees before and after the QC(Quality Control) activities were compared and showed the effects of the QC activities[3][7].

The lead time for customers is one of indicators that show the quality of the service. It basically means cooking time of the meal which a customer ordered. In this paper, the lead time refers the time from the end of taking order to the start of serving shown in Fig.1. It would be best as shortly as possible for customers. Yang et al.[8] investigated the length of waiting time and mood states moderation in a restaurant or a dental hospital. If a number of orders is larger than a capacity for work of employees, the act of dealing with customers and cooking meals take more time so that an average time that customers wait for their meal is getting longer. To survey the lead time for customers leads to clarifying a pressure of the work and improvement employees' behavior. Our goal is to realize lead time estimation.

III. Waiting Operation Detection

A. Overview of WOD

Waiting operation detection(WOD) is a method to estimate when an employee did services to customers. WOD has two purposes. First, it is preprocessing for LTE. We must find when the employee did two kinds of operations, *take order* and *serve*, before LTE. It also can be applied a service operation estimation technique[9]. Second, it enable us to know a pressure of the employee. We think waiting operations are harder burden than non-waiting operations because an employee takes care of customers. The waiting operation(WO) means providing a service for customers in a restaurant. For example, it contains to welcome guests, conduct to seats, take order, serve a meal and take accounts. On the other hands, the non-waiting operation includes to cook a meal, prepare a drink, deliver(before serve) and cleaning a guest room.

Fig.2 shows the average of speech ratio of two employees during their operations. Four operations from the left, waiting operations, indicate high speech ratio. The other operations, *deliver* and *clean*, are lower speech ratio than the waiting operations. It describes the waiting operations have more burden than other operations. Moreover, the result helps to classify waiting or non-waiting operation. The two speakers have big differences in *serve* and *clean* operations. When each

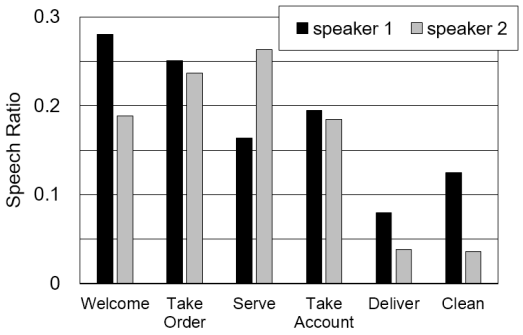

Fig. 2. Speech ratio of two employees during their operation.

speaker serves meals, the speech ratio of speaker 2 is higher than that of speaker 1. Because speaker 2 often explained about meals for customers after putting them. When each speaker cleans a guest room, the speech ratio of speaker 1 is higher. That means speaker 1 actively removed used dishes from the table after the customers had finished eating. In this way, the employees' works reflect the difference of the speech ratio.

B. Flow of WOD

There are 3 features in the waiting operations. An employee deals with customers while (1)*talking* to them. They are (2)*staying* in a (3)*guest room*. Fig.3 shows the flow of WOD. We extract three features of WOD by sound data and locational data. Speaking score $S_V(t)$ is calculated by the result of VAD. First, we calculate speech ratio of each frame. Then, before and after several frames are weighted and added to $S_V(t)$. Finally, $S_V(t)$ is normalized so as to be within the range from 0 to 1. An employee doesn't speak always during WO. $S_V(t)$ is intermittent. By smoothing before and after several frames, $S_V(t)$ changes gradually.

Staying score $S_M(t)$ is calculated by locational data. If an employee moves largely like crossing over the boundary of two areas, $S_M(t)$ becomes lower. If an employee stays one area and guest room, $S_M(t)$ becomes higher. $S_M(t)$ is the inverse of the variance of employee's position coordinate data between several frames. Then, $S_M(t)$ is normalized as with $S_V(t)$.

Formula (1) is an integration formula of $S_V(t)$ and $S_M(t)$. $S_{serve}(t)$ represents a likelihood of a waiting operation. The weight factors λ are changed by 0.1 to implement formula (2).

$$S_{serve}(t) = \lambda_V S_V(t) + \lambda_M S_M(t) \quad (1)$$

$$\lambda_V + \lambda_M = 1 \quad (2)$$

We detect WO of the employee by thresholding $S_{serve}(t)$. If the employee was not in guest room, we forcibly regard the operation of frame t as a non-waiting operation.

C. Experiment

We conducted the experiment to discriminate when and whether the employee did WO or not. Table I shows the

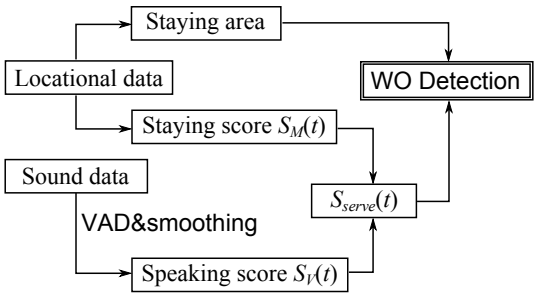

Fig. 3. Flow of WOD Processing.

TABLE I
EXPERIMENTAL CONDITIONS OF WOD.

#subjects	3 employees
data length	17h 2m 15s
#frames	12267
#WO	288 times
#WO frames	1545
threshold	0.5

experimental condition of WOD. We evaluated the method described in Sec. III-B for the behavior data of three employees. The length of each frame is 5 seconds. We test four types of smoothing of the speaking score, no smoothing, before and after 1 frame(15 sec.), 2 frames(25 sec.) and 3 frames(35 sec.).

We used F-measure as the evaluation measure. F-measure is a harmonic mean between the precision and the recall of WOD. The precision is the rate of correct WO frames to estimated WO frames. The recall is the rate of correct WO frames to truth WO frames.

D. Result

Fig.4 shows the comparison result of WOD between four types smoothing. By smoothing the speaking score, each accuracy is all improved. The accuracy when the ratio of the speaking score is 0.7 and smoothed before and after 2 frames is best. Integration of sound data and locational data needs smoothing since the variations of these data are difference largely. Fig.5 shows the accuracy of WOD by a unit of WO. When the ratio of the speaking score is 0.9, the accuracy is 87%. Like frame-level detection, the speaking score is more important than the staying score. However, the staying area data enable to reject about 30% of miss-estimated WO. On the other hand, ratio of the staying score is higher than that of the speaking score. The reason is although the employee was stopping, locational data sometimes described greatly variation so that the staying score is getting larger. It may cause the locational data was sometimes rectified excessively. If the accuracy to estimate the locational data were improved, multi-data processing get worth more.

IV. LEAD TIME ESTIMATION

The lead time is an index of service qualities in a restaurant. It shows the time between ordering a meal and serving. The time would be best as quickly as possible for customers. There

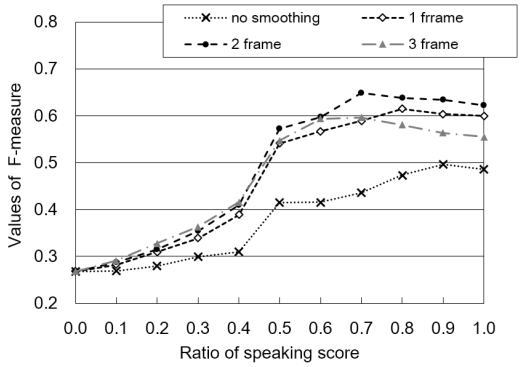

Fig. 4. Comparing the accuracies of WOD by smoothing.

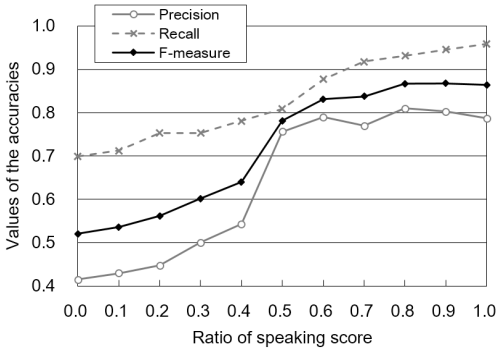

Fig. 5. Comparing the accuracies of WOD by a unit of WO.

are several ways to estimate the lead time for customers. An approximate cooking time is depend on a meal. The name of the ordered meal is recorded to accounting data. We may be able to know rough lead time by using accounting data. However, a pressure of the work changes depending on the number of customers and orders. It makes a difference of the lead time even same orders.

The other approach is to estimate taking order operation and serving operation of the employee for a group of customers. The time from the end of taking order to the start of serving is the lead time for customers. This approach is better than the previous approach. Because an actual lead time for customers is longer than the cooking time. Therefore, we can know the quality of service more correctly. However, there are two issue to this approach. First is the accuracy of these operations estimation. Second is how to map these operations to customers.

This paper shows the solution to realize LTE by estimating the two operations from behavior data of the employee and accounting data. First, we describe how to estimate *take order* operations and report the examination. Second, we notice the important points to estimate *serve* operations.

A. Taking order operation

Fig.6 shows the description of the relation between taking operations by employees and sound and account data. Sound data contains the sound of POS terminal after an employee

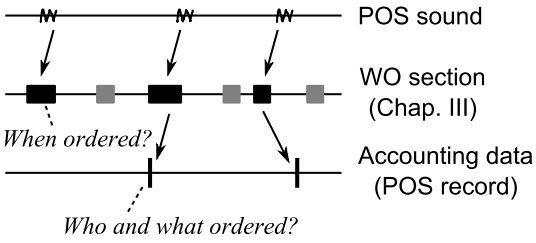

Fig. 6. Relation between *take order* and sound/POS data.

took order. We can estimate the time to take order correctly with combination of POS terminal sound and the WOE result. Moreover, accounting data has the time which an employee took order in unit of a minute, which customers ordered and which dishes were ordered. It enables to associate each taking order operation with customers and dishes. The accuracy of estimated taking order operations were over 95%.

Taking the opposite point of view, there were twice as much opportunities for operations of the POS terminal as the times recorded to the POS terminal. For example, the employees took additional orders and served free drinks. It is interesting that we may be able to notice the other business that we didn't measured.

B. Serving operation

When an employee serves a meal, he or she doesn't use POS terminal. Therefore, sound data and locational data are required to estimate serving operations of the employee. If speech environment of sound data is clean, we can use a speech recognition technique for some keywords related to taking order and serving operations. But our sound data has many kinds of noises that prevent to recognize the keywords. Besides each employee uses different words even if a situation is same. Therefore, we should consider other methods in order to estimate serving operations.

We manually labeled the time when the employee serves the meals. Fig.7 shows the reason of judgment to determine true *serve* operations from sound data of the employee. The key phrases often used in *serve* operations, "Here you go." and "Excuse me." were also used in other kinds of operations. The specific key phrases , "Please be careful ..." and "Is this everything?" were used rarely. So we judged *serve* operation by the name of the meal, the explanation of the meal and the sound of dishes. When we estimate serving operations, distribution of the speech ratio on one WO, a number of changing the speakers, speaker identification, putting down sound of dishes are considered as the features.

C. Toward lead time estimation

Fig.8 shows the start of the lead time based on sound and WO data. We can estimate taking order operations. Thus, we can find the start of the lead time automatically. However, we must find the end of the lead time for LTE. To estimate serving operations and associate serving operations with orders are our future task.

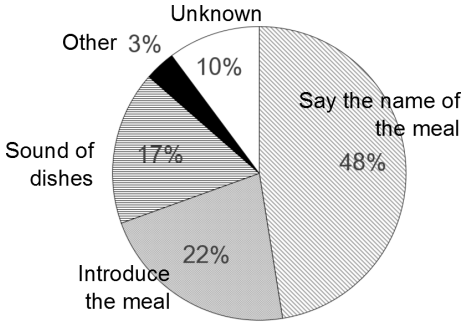

Fig. 7. The types of *serve* operations by labeling manually.

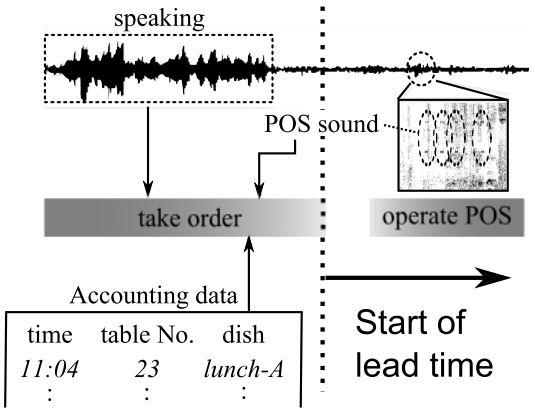

Fig. 8. What happens in the start of a lead time?

V. Conclusions

This paper showed the role of sound data for WOD of employees in a restaurant. The speech ratio between the waiting operation and the non-waiting operation has the large differences. Working styles of the employees were reflected to the speech ratio so that it is considered to be useful to introduction and education for employees.

In WOD, sound data and locational data of the employees were integrated. By smoothing the speech ratio, the accuracy of WOD is improved. If the result of WOD allows a little error, it can make use of LTE. In LTE, the sound of POS terminal is useful for estimate taking order operations.

In future works, we should identify serving operations from WO and complete the lead time estimation. Moreover, frame-level WOD will require greater precision and apply to service operation estimation techniques.

Acknowledgment

The authors thank Ganko Food Service Co., Ltd., and staff in the Ganko Ginza 4-chome restaurant for their cooperation with our experiment.

References

[1] B.Hartmann et al., "Identifying Redundancy and Exposing Provenance in Crowdsourced Data Analysis," *IEEE trans. on Visualization and Computer Graphics*, 2013.

[2] H.Noma et al., “Wearable sensors for auto-event-recording on medical nursing-user study of ergonomic design,”in *Wearable Computers, 2004. ISWC 2004. Eighth International Symposium on. vol.1.*, 2004.
[3] T.Fukuhara et al., “Measuring and evaluating real service operations with human-behavior sensing: a case study in a Japanese cuisine restaurant,” *in Proc.FCV2013*, 2013.
[4] M.Takehara et al., “The role of speech technology in service operation estimation,” *in Proc.O-COCOSDA2011*, 2011.
[5] Y.Guo et al., ”Robust voice activity detection based on adaptive sub-band energy sequence analysis and harmonic detection,” *in Proc.Interspeech2007*, 2007.
[6] T.Kawase et al., ”Improvement of utterance clustering by using employees’ sound and area data,” *in Proc.ICASSP2014*, 2014.
[7] M.Takehara et al., ”Measurement and analysis of speech data toward improving service in restaurant,” *in Proc.O-COCOSDA2013*, 2013.
[8] C.Yang et al., ”Event Values on the Length of Waiting Time or the Waiting Time Information,” *in Proc.APPEEC2011*, 2011.
[9] R.Tenmoku et al., ”Service-operation estimation in a Japanese restaurant using multi-sensor and POS data,” *in Proc.APMS2011*, 2011.