# A WEB-BASED MULTILINGUAL UTTERANCE COLLECTION SYSTEM FOR THE MEDICAL FIELD

Takashi Yoshino, Taku Fukushima and Ryuichi Nisimura
*Faculty of Systems Engineering, Wakayama University, 930 Sakaedani, Wakayama, Japan*

Keywords: Voice collection, Intercultural communication, Parallel texts, Medical field, Web-based system.

Abstract: We have developed a web-based multilingual utterance collection system, named OTOCKER, for the medical field. The purpose of OTOCKER is to act as a voice data collection platform for intercultural communication. Although speech synthesis systems have improved significantly, fluent and smooth speech synthesis is still a problem. Currently, it is difficult to synthesize speech in different languages. In the medical field, in particular, it is important that the intended meaning of spoken words be conveyed effectively along with the different nuances. Therefore, we use an utterance collection system that collects peoplefs voices directly. The limitations of this system are (1) an insufficient number of correct sentences pertaining to the medical field and (2) that easy participation in both voice record and collection are difficult. We can solve the first problem by using a system that collects parallel medical texts. The second problem can be solved by using w3voice — web-based voice-recording system. This system can run only on a web browser. This paper presents the design of OTOCKER, its prototype, and the results of its trial.



## 1 INTRODUCTION

Recently, there has been a sharp rise in the number of foreigners visiting Japan. Moreover, the number of foreign residents in Japan is increasing steadily. The Immigration Service reported that approximately 6.73 million foreigners visited Japan in fiscal year 2006. That is, there was a 10.0% increase as compared to the previous year. The Immigration Service also reported that the number of foreign residents in fiscal year 2006 was approximately 2.08 million, which was a 3.6% increase as compared to the previous year. However, many of these foreigners who either live in or visit Japan cannot speak Japanese. Further, there are some people who can speak a language but do not know how to read and write in that language. Therefore, translation and voice support has to be provided to such people. For instance, foreigners are now provided such voice and translation support in the field of medicine by a medical treatment interpreter. The interpreter support is also provided over the telephone. This telephone interpreter support service is available in seven languages and is used by more than 4000 people a year. The number of languages supported and the considerable number of users place an excessive load on the system. Moreover, a further increase in the load is expected in the future.

This paper presents a Web-based multilingual utterance collection system, named OTOCKER, which can collect multilingual voice data related to the field of medicine. We have developed OTOCKER and carried out a trial experiment using this system. In this paper, we will discuss the effectiveness of this system.

## 2 RELATED WORK

There are some researches to collect speech corpus(Maekawa et al., 2000; Fujii et al., 2008; Griol et al., 2008). Maekawa et al. collect a spontaneous speech corpus of Japanese to make it useful for natural language processing and phonetic or linguistic studies(Maekawa et al., 2000). Fujii et al. collect a speech corpus of Japanese classroom lecture to extract an importance sentences(Fujii et al., 2008). David et al. propose two speech corpora acquisition techniques(Griol et al., 2008). There is no research to collect multilingual voice data for being provided to other voice-based systems.

Considerable research has been carried out in order to support intercultural communication via machine translation(Ishida, 2006; Yoshino et al., 2008).



Yoshino T., Nishimura R. and Fukushima T.
A WEB-BASED MULTILINGUAL UTTERANCE COLLECTION SYSTEM FOR THE MEDICAL FIELD.
DOI: 10.5220/0001844503700375
In *Proceedings of the Fifth International Conference on Web Information Systems and Technologies (WEBIST 2009)*, page
ISBN: 978-989-8111-81-4

However, the accuracy of machine translation is not sufficient for it to be used in the field of medicine. This is because any mistake in communication can directly affect a patient's life.

Therefore, multilingual corpora, consisting of text related to the medical field, are collected by a translation system for parallel translation(Yoshino et al., 2009). The following three-step process is followed by this system: (1) The necessary corpus is registered with the system by the user. (2) The corpus is translated into different languages by other users. (3) The translation is clubbed with the original corpus.

The collected corpora are then made available to other medical support systems. As of now, this translation system is available in five languages (Japanese, English, Chinese, Korean, and Portuguese).

However, this translation system cannot provide support to a foreigner who is unable to read any of the languages the system is available in. In such a case, it is necessary to provide voice support. Different voices have been collected for speech translation research before. However, the use of these collected voices has become one of the most challenging issues in the field since not all voices can be understood by everyone.

Therefore, the purpose of this research is to collect accurate voices for providing voice translation support in the field of medicine(Miyabe et al., 2007).

## 3 ATTRIBUTES OF UTTERER

Voice data collected in the utterance collection system are meant to be provided to other voice-based systems. Therefore, the attributes of the person whose utterances are recorded (the utterer) need to be known. This system considers the following four attributes of an utterer.

1. **Sex and Date of Birth.**
   This system records the sex and the date of birth of the utterer since prior to using the recorded data, it is necessary to know whether the utterer is a male or a female and whether he/she is an adult or a child.
2. **Native Language.**
   Since a person's native language affects his/her pronunciation, the system also records the native language of the utterer.
3. **Proficiency in Nonnative Language.**
   It is necessary to know whether the utterer is comfortable with the nonnative language in which he/she is recording. Therefore, we asked the person to rate his/her proficiency in the nonnative language he/she was prepared to record in. The definitions of the levels of proficiency are given in Table 1.
4. **Dialect.**
   The different ways in which a language is spoken in different countries or in different regions of the same country are called dialects. Since the quality of voice data is also dependent on these differences, the system even maintains a record of the dialect spoken by the utterer. This system classifies the variants of a language into the following two types:
   - Variants in which both pronunciations and characters are different
   - Variants in which words are spelt in almost the same way but their pronunciation is different

   Chinese and Portuguese are examples of the first type of classification. Chinese has two dialects, Pekingese and Cantonese. The Portuguese spoken in Portugal is different from the Portuguese spoken in Brazil. These dialects are mutually different in terms of both pronunciation and spelling. In particular, the pronunciation of both the Chinese dialects is so different that people speaking the two different dialects cannot understand each other. English is an example for the second type of classification. Some words in American English, British English, and Australian English are almost similar in spelling but differ in pronunciation.
   This system refers to the language variants that fall in the latter category as "dialects," and the ones that fall in the former category are classified as different languages.

## 4 DESIGN OF OTOCKER

This section describes the design of a Web-based multilingual utterance collection system called OTOCKER. We used PHP as the development language. In this system, the voice is registered and replayed on a Web browser.

### 4.1 User Management

This system defines four types of roles for users: manager, voice editor, voice registrant, and voice replay user.

Table 2 shows the system rights for each role. The system limits the functions of each role in order to manage the large number of users.

Table 1: Description of four proficiency levels.

| Proficiency levels | Description |
|---|---|
| 1 | I can convey my opinions and ideas clearly in this language. |
| 2 | I can manage day-to-day conversation in this language. |
| 3 | I am a poor speaker of this language. |
| 4 | I do not speak this language. |

Table 2: System rights for each role.

| Rights | Roles | | | |
|---|---|---|---|---|
| | Manager | Voice editor | Voice registrant | Voice replay user |
| Add user, edit user, delete user | O | X | X | X |
| Delete voice data | O | O | X | X |
| Delete my voice addition | O | O | O | X |
| Add new voice data | O | O | O | X |
| Replay voice data | O | O | O | O |

*A voice registrant can delete his/her own registered voice.

"O" shows that the role has the rights. "X" shows that the role does not have the rights.

## 4.2 System Configuration

Figure 1 shows the system configuration. Figure 2 shows the screen transition chart. In order to run correctly, this system needs a Java-enabled Web browser. The system uses w3voice(Nisimura et al., 2008) for collecting voice data on the Web browser. w3voice can operate on Windows, MacOSX, and Linux; however, Java has to be installed on these operating systems. The voice of an utterer can be recorded only by a mouse click. We use w3voice to reduce the users' threshold.

## 4.3 Voice Data Registration and Replay

Figure 3 shows the flow of voice data registration and replay.

### 4.3.1 Flow of Voice Registration

1. The system presents the corpus list obtained from the database of a medical-field corpora-sharing system to a voice registrant.
2. The voice registrant selects a corpus from the list.
3. The user utters the corpus into the microphone. Figure 4 shows the voice registration screen. After recording the voice data with a voice recorder, the system saves this voice data into a directory. Then, the attributes and information of the voice of the utterer are added to the database.

### 4.3.2 Flow of Voice Replay

1. The system presents the corpus list obtained from the database of the medical field corpora sharing system to a voice player.
2. The voice player selects a corpus from the list.
3. The voice list for the corpus is displayed.
4. The user selects the voice to replay.
5. The system obtains the voice file name from the database and replays the voice. Figure 5 shows a screenshot for a voice replay. When the user clicks the voice replay button, the voice replay begins. During the replay, the system shows the profile of the utterer to the voice replay user.

# 5 TRIAL EXPERIMENT

We carried out a trial experiment using the proposed system. The purpose of the experiment was to evaluate the usability of OTOCKER.

## 5.1 Examinees

Table 3 shows the examinees' native language and the other languages that they can speak. Table 3 provides the language-based distribution of the examinees. The examinees are volunteer interpreters.

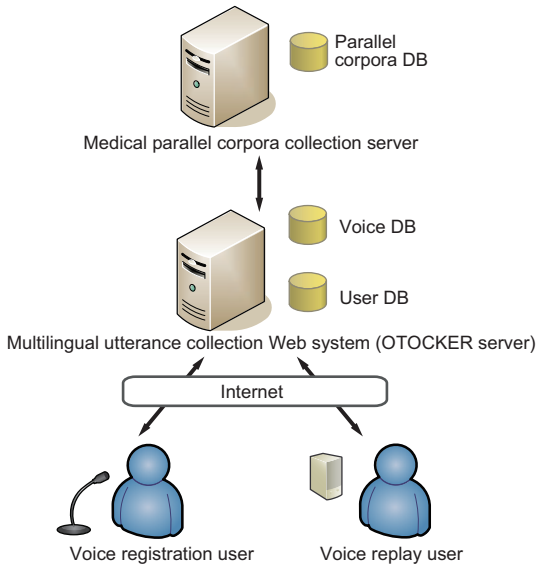


Figure 1: System configuration of OTOCKER.

## 5.2 Experimental Procedure

We carried out the following procedure:

1. We explained the background and the purpose of this system to the examinees.
2. Examinees registered their user information.
3. We explained the usage of the system.
4. The examinees recorded their own voice.
5. We asked the examinees to fill out a questionnaire.

# 6 RESULT AND DISCUSSION

## 6.1 Overall Expression of OTOCKER

Table 4 shows the results of the questionnaire survey. The survey was based on a description-type questionnaire, and for the evaluation of the survey results, we used a five-item Likert scale.

1. Item 1: I was able to register my voice easily.
   The average value of the result is 4.2 on the five-item evaluation. This value implies that even though the examinee was not accustomed to the operation of a PC, he/she could operate the system easily. Some of the examinees also expressed surprise at the fact that their voice could be recorded so easily and clearly just by clicking on a button.
2. Item 2: This system is easy to use.
   The average value of the result is 3.6 on the five-item evaluation. In this case, the examinees were divided in their opinions. Some people said that the operation of this system was easy or very simple. The others said that they required a guide in order to use this system. Therefore, we have concluded that the usability of the system needs to be improved further.

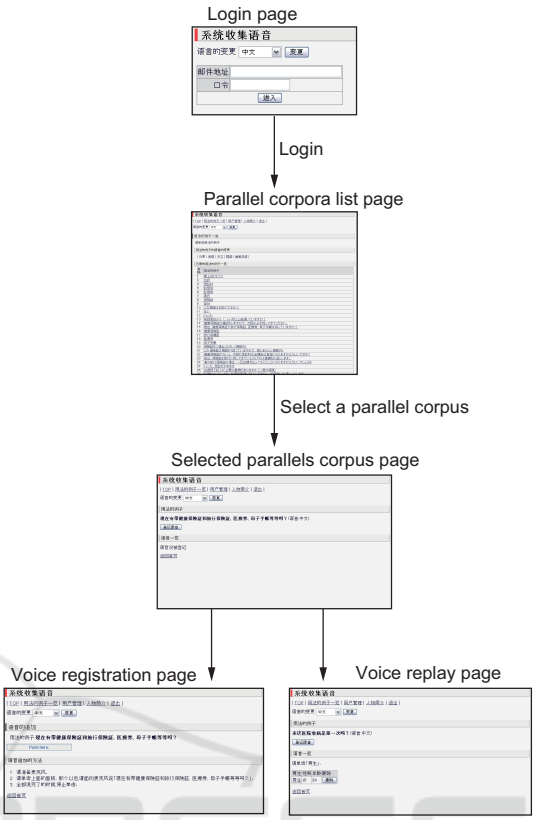


Figure 2: Screen transitions in OTOCKER.

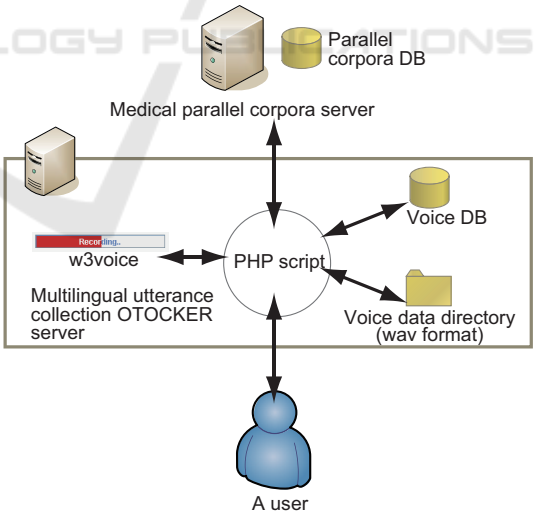


Figure 3: Flowchart of voice registration and replay.

Table 3: Language-based distribution of examinees.

| | |
|---|---|
| Native language | Japanese: 4 persons; Portuguese: 1 person |
| Other languages that can be spoken | Japanese: 5 persons; English: 2 persons; Portuguese: 2 persons; Spanish: 2 persons |

Table 4: Results of questionnaire survey.

| No. | Items | Evaluation value | | | | | Average |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | |
| 1 | I was able to register my voice easily. | – | 1 | – | 1 | 3 | 4.2 |
| 2 | This system is easy to use. | – | – | 3 | 1 | 1 | 3.6 |
| 3 | I want to register my voice on this system. | 1 | – | – | 3 | 1 | 3.6 |
| 4 | I want to use this system continuously. | 1 | – | 2 | 1 | 1 | 3.2 |
| 5 | I would not like my voice to be used by other systems. | 2 | 2 | 1 | – | – | 1.8 |

1) Evaluation value: 1: Strongly disagree, 2: Disagree, 3: Neutral, 4: Agree, 5: Strongly agree

2) The numbers in each evaluation value column denote the number of people who held the view reflected by that particular evaluation value.

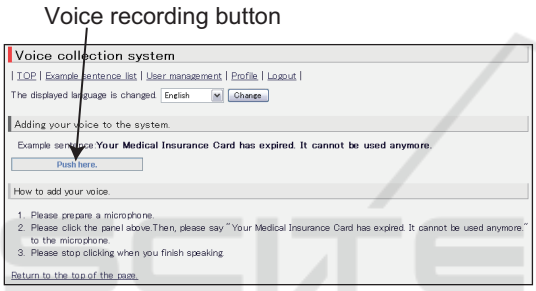


Figure 4: Voice registration screen.

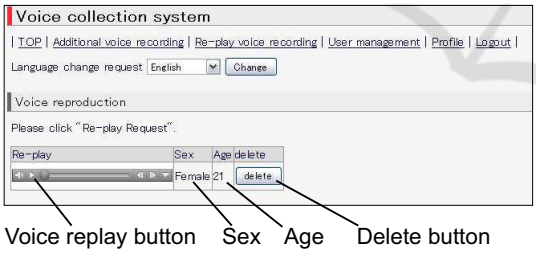


Figure 5: Voice replay screen.

3. Item 3: I want to register my voice on this system, and Item 4: I want to use this system continuously. The average values for Items 3 and 4 are 3.6 and 3.2, respectively. As responses to these items, we received positive comments such as "I think that the system can help many foreigners if it is used in hospitals across Japan," "I would be glad to register my voice on this system," and "I am happy to have registered my voice with this system." However, we also received few negative comments such as "I was not happy with the quality of recording" and "I find it disturbing that my voice will be used in a hospital."
4. Item 5: I would not like my voice to be used by other systems.
   The average value of the result is 1.8 on the five-item evaluation. This result shows that there was little resistance to the use of the voice. We did not expect this result and believe that it could be attributed to the fact that the examinees were volunteers.

## 7 CONCLUSIONS

We designed and developed a Web-based multilingual utterance collection system, named OTOCKER. The purpose of OTOCKER is to act as a voice collection platform for intercultural communication. In this paper, we presented the configuration of this system and the results of trial experiments.

## ACKNOWLEDGEMENTS

This work was supported by SCOPE (Strategic Information and Communications R&D Promotion Programme) of Ministry of Internal Affairs and Communications.

## REFERENCES

Fujii, Y., Yamamoto, K., Kitaoka, N., and Nakagawa, S. (2008). Class lecture summarization taking into account consectiveness of important sentences. In *INTERSPEECH2008*, pages 2438–2491.

Griol, D., Hurtado, L. F., Segarra, E., and Sanchis, E. (2008). Acquisition and evaluation of a dialog corpus through woz and dialog simulation techniques. In (ELRA), E. L. R. A., editor, *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco.

Ishida, T. (2006). Language grid: An infrastructure forintercultural collaboration. In *SAINT-06, IEEE/IPSJ Symposium on Applications and the Internet*, pages 96–100. IEEE Press.

Maekawa, K., Koiso, H., Furui, S., and Isahara, H. (2000). Spontaneous speech corpus of japanese. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation (LREC2000)*, pages 947–952.

Miyabe, M., Fujii, K., Shigenobu, T., and Yoshino, T. (2007). Parallel-text based support system for intercultural communication at medical receptions. In *Proceedings of The First International Workshop on Intercultural Collaboration (IWIC2007)*, pages 182–192.

Nisimura, R., Miyake, J., Kawahara, H., and Irino, T. (2008). Speech-to-text input method for web system using javascript. In *SLT2008, IEEE Workshop on Spoken Language Technology*.

Yoshino, T., Fujii, K., and Shigenobu, T. (2008). Availability of web information for intercultural communication. In *PRICAI-08, The Tenth Pacific Rim International Conference on Artificial Intelligence*.

Yoshino, T., Fukushima, T., Miyabe, M., and Shigeno, A. (2009). A web-based multilingual parallel corpus collection system for the medical field. In *IWIC-09, ACM Workshop on Intercultural Collaboration*.