
# Classrooms and voice recognition applications in a foreign language teaching

Yuichi ONO[1]; Manabu ISHIHARA[2]; Akio ONISHI[3]; Mitsuo YAMASHIRO[4]

[1] University of Tsukuba, Japan

[2] Oyama National College of Technology, Japan

[3] Ashikaga Institute of Technology, Japan

[4] Version2, Japan

## ABSTRACT

The advantages of using Voice recognition software in a foreign language course to enhance speaking accuracy and fluency have been suggested in the literature (1,2). One of the difficulties faced with Voice recognition software is a "noise" around the speakers. In our campus, the normal class size is around 40 for one English course. This means is that the software has to recognize speaker's voice correctly among other speakers' "noise". In this study, we carried out several experiments in terms of the system and the devices of the classroom. We picked up two kinds of classroom. One is Computer-Assisted Language Learning (CALL) room, a specialized classroom for foreign language learning, where there are wired individual PC and headphone with microphone for each student. The other classroom is a normal classroom, where tablet computers or mobile devices were utilized. Our expectation is that the rate of recognition should be better in CALL classrooms than that of normal classroom with tablet computers, since students use headphones and microphones in a CALL classroom. The results of our study indicated that this hypothesis was true only with sight significance. This paper suggests that it is still difficult to integrate Voice recognition software into normal classroom with local application with tablet PC.

Keywords: Building Noise Control, Classroom Noise, Voice recognition, Pronunciation Training

## 1. INTRODUCTION

There are numbers of applications for foreign language education uses. They can teach proper pronunciation, in addition to helping a person develop fluency with their speaking skills. Since the use of Tablet has become widely common, a lot of research on educational use of technology has suggested that these applications can be used outside the classroom situations, anywhere and anytime learners would like. Recent trends in Japanese Teaching English as a Foreign Language (TEFL) settings are the use of software like Text-to-Speech, Speech-to-Text, Pronunciation Training, and so on.

In spite of the conveniences and benefits provided by the software, there is a big issue in using this type of software in a large-scale classroom. The classroom is surrounded with various kinds of noise, which seriously affects the accuracy of natural Voice recognition by software. In Japan, it is still common that the number of the class members is more than 30. In the case of the first author, the average number of class members is 38.5 (Min.:33, Max.:47). In the class, for example, when all the members started loud speaking practice simultaneously, the sound pressure level is more than 70 dB, which is a really tough situation for the software.

[1] ono.yuichi.ga@u.tsukuba.ac.jp
[2] ishihara@oyama-ct.ac.jp
[3] yamasiro@ashitec.ac.jp
[4] a-ohnishi@ver2.jp

Some previous studies have suggested the usefulness in the integration of Voice recognition software into course curriculum, resulting in pedagogical effect in speaking and other linguistic proficiencies. However, looking closely at the situation, the class-size tends to be small, less than 30. Moreover, the class is conducted in a special room for Computer-Assisted Language Learning (CALL), which is equipped with a special headphone and microphone, reducing the effects by classroom noise.

The purpose of this paper is to examine how the software works under different situations: the use of CALL equipment in a CALL room and Tablet-PC use in a classroom. In our experiments, we considered two situations; (i) the use of web application with a CALL equipment, and (ii) the use of local application installed in the computer (Tablet PC). The result of our experiments will show that it is still difficult under the current situation that learners work with Voice recognition software satisfactorily at the same time in the classroom regardless of type of the classrooms and location of the process.

## 2. BACKGROUNDS

### 2.1 Voice recognition Software

According to the location of the software engine, the software can be regarded as consisting of two types; on-line type and off-line type. If the engine is installed in the server on the web, it can be referred to as “Web Application”. If it is installed in end-user’s PC or Tablet, it is referred to as “Local Application”. It is generally agreed that the software in off-line use operates quickly and more accurately than that in on-line use, since web-application is subject to various kinds of noise through the internet access to the server, causing the delay. A large amount of information operated by a number of users simultaneously sometimes results in failure. But the use of Web application provides a better efficient learning environment, because it can be easily incorporated into the Learning Management System (LMS) on the web. Installing local application on each individual PC costs a lot of money, and the instructor has to control each software used by every user’s computer, which causes the instructor a great amount of troublesome especially when the instructor is not familiar with information and computer literacy.

The software we employed for this experiment is “Hatsuon-Kentei (Pronunciation Test)” by PRONTEST, a Japanese company which created a pronunciation courseware under the collaboration with National Institute of Advanced Industrial Science and Technology (AIST). The courseware is composed of several speaking tasks; pre-reading/recording, practice, model pronunciation, prontest (pronunciation test), feedback of the test result, post-reading/recording, and reflection. The screen shots of parts of the courseware are given in Figures 1 and 2. The LMS server where the application works is located in our campus.


Figure 1 – Screenshot (Scoring)

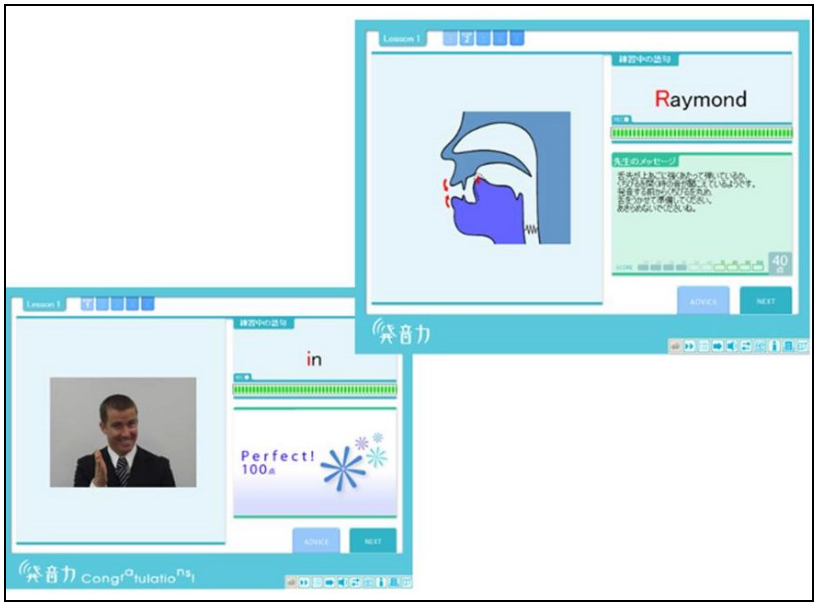



Figure 2 – Screenshot (Visualized feedback)

The use of CALL room makes it easier to get access to the server, because the location is the closest, actually both are inside the same building. The server to be installed was GLEXA; a Learning Management System for Foreign Language Learning by Version 2, Japan. Installed in the LMS, the application works as a teaching material or course material to be used in the classroom or outside of it. The experiment was conducted in a CALL classroom and a Tablet-CALL classroom in our institution. The hardware outline is illustrated in Figure 1 below, where the location of the web and local applications is specified.

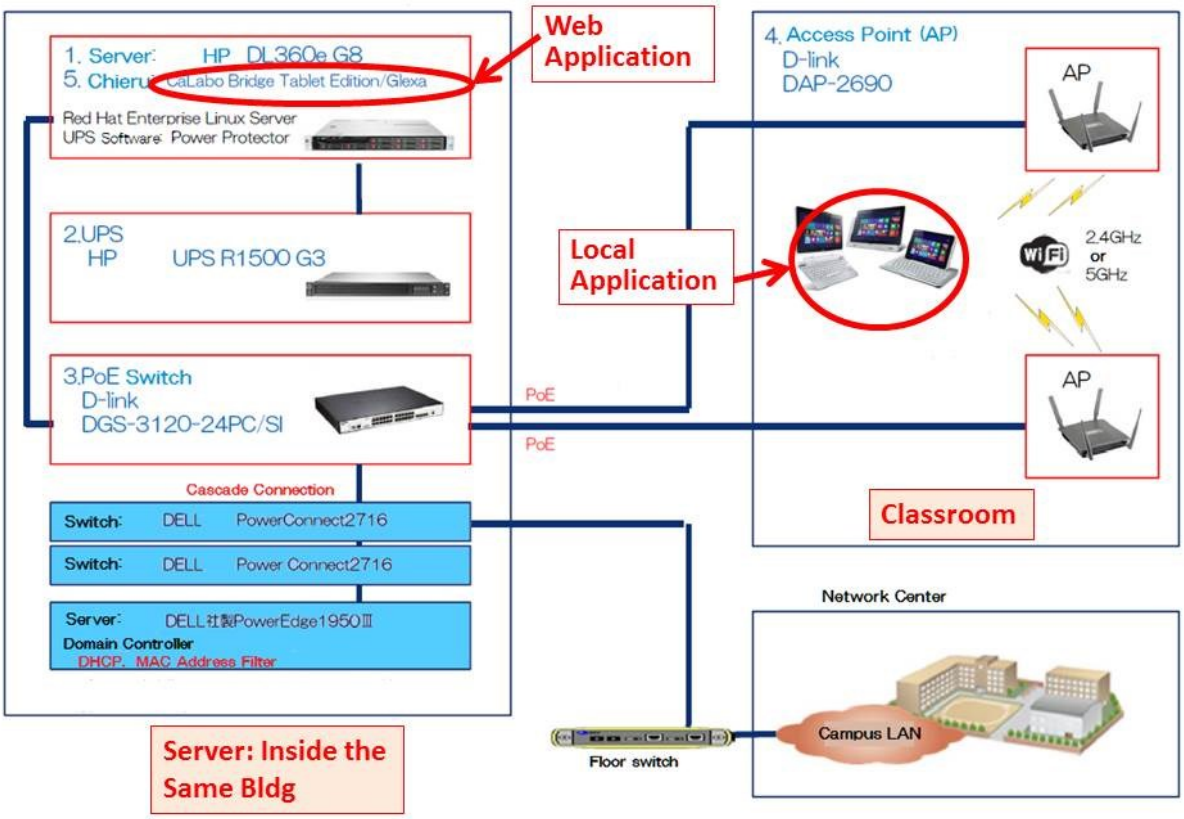



Figure 3 – Hardware outline

### 2.2 Classroom Environments

Needless to say, noise has a negative effect on language learning. Considering various tasks in the classroom, listening and speaking activity is greatly affected by noise in the classroom. The same is true of Voice recognition by machine. Noise greatly affects the accuracy and operability of the application. In order to protect learners from being disturbed by noise, the classroom for language learning is equipped with headphone with microphone, as shown in Figure 4. It can be said that in a CALL room, students can work on speaking and listening activity more comfortably. But in a normal non-wired classroom where there is no individual PC or headphone is available, the system had to endure the difficulty created by noise and had to find a solution only with Tablet PC.

Figure 4 – Use of Web application with CALL equipment

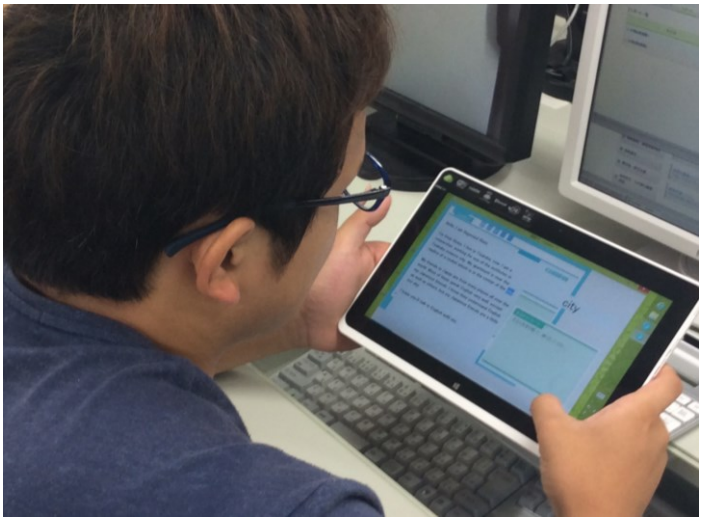


Figure 5 – Use of local application on tablet PC

### 2.3 Questions to be asked

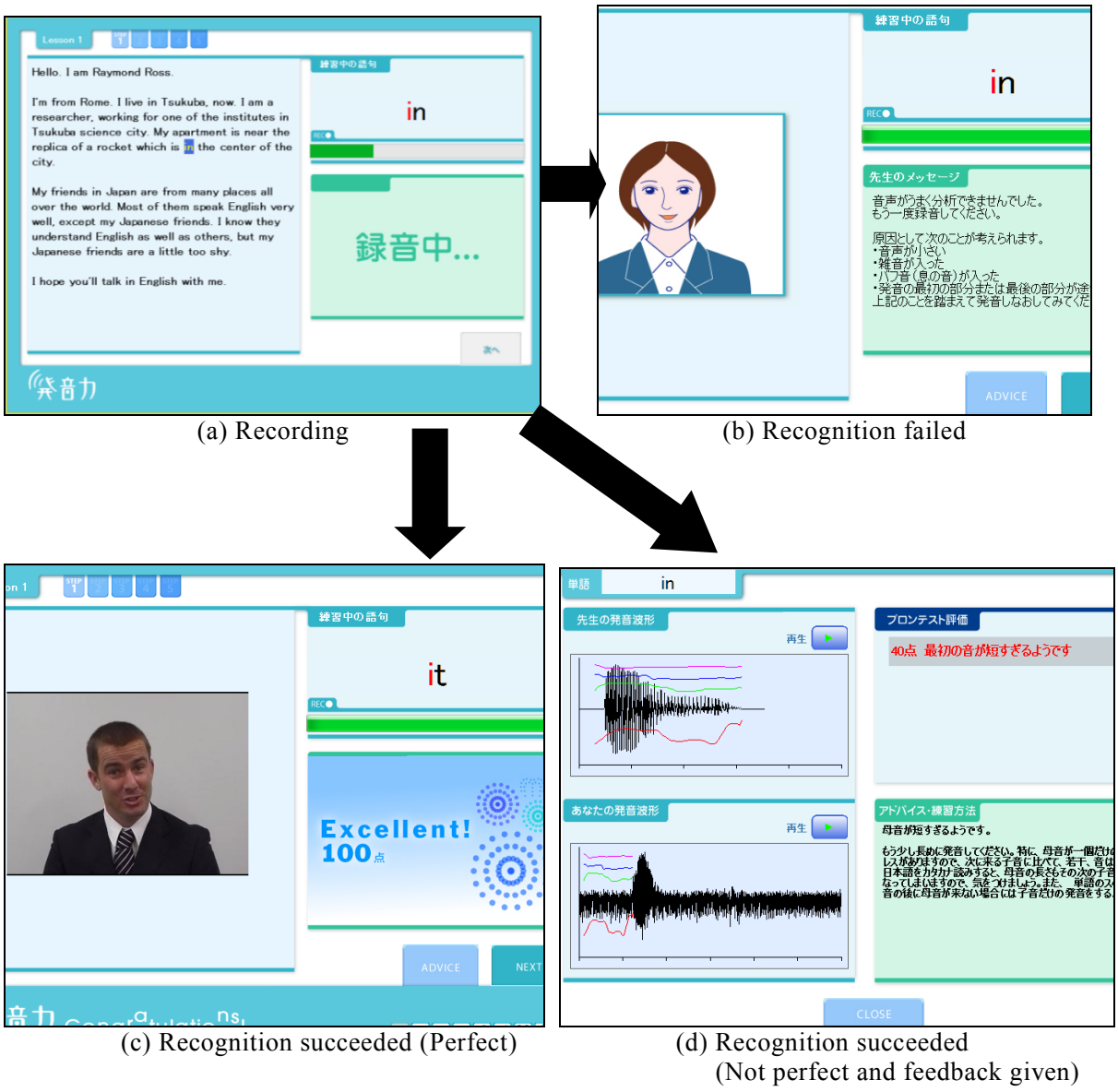
Keeping what has been said in mind, this paper as a pilot study attempts to examine the practicality of using Voice recognition software as a simultaneous speaking practice in a classroom at the same time. Since the use of local application in a CALL classroom was validated in some previous studies (2), this paper set up the following two situations to advance the use of software in a large-size classroom:

Local application on a tablet PC in a CALL room
Web application in a CALL classroom

In our experiment, we used the same CALL classroom in order to prevent noise variability caused by using different classrooms. In other words, classroom factors are controlled in this study and similar tasks are assigned, creating the situation that the level of noise is almost similar in both settings.

## 3. Experiments

45 students participated in the study. The class was conducted in the CALL classroom as shown in Figure 4. They were asked to report on how well the application succeeded in recognizing the speech. The author instructed the students to judge whether the software succeeded in recognition according to the following standards. The courseware proceeds to "prontest" (Pronunciation test) session. The student speak the word to be tested (a). If the software does not recognize and produce the assessment, the recognition is judged as "failed". As in (b). If it succeeded in recognizing, the software gives the "prefect" score as in (c) or provides the feedback when imperfect on what is to be required as in (d). In both cases, the software recognized the student's pronunciation; thus judged as "succeeded".




(a) Recording

(b) Recognition failed

(c) Recognition succeeded (Perfect)

(d) Recognition succeeded (Not perfect and feedback given)

Figure 6 – Recognition standard: (b) failed; (c) and (d) succeeded

### 3.1 Local application on wired desktop PC

The local use of the software CALL equipment was demonstrated in a college course (2), where no trouble or hardware and software problems were reported. The improvement in terms of speech duration, speech power, F0 (pitch), the ratio of vowel and consonant length and power was observed and the results of the popular computerized assessment system for English Communication (CASEC)

showed that their overall improvement was observed. Thus, we assume that the local use of the software with CALL equipment in a CALL room is reliable. Under this assumption we will have two experiments under different situations; the use of web application with CALL equipment and the use of local application with Tablet PC. Since we also assume that the use of web application with Tablet PC is least reliable considering the power of the machine and instability of non-wired communications. Under these things, our expectation is as follows:

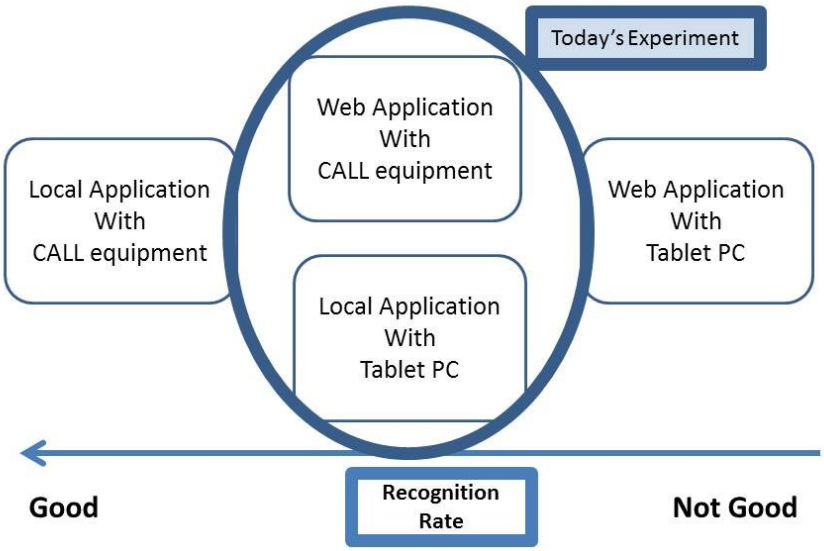



Figure 7– Our expectation

### 3.2 Local application with tablet PC

The prontest software is installed in tablet PC (Figure 5). The operating system is windows 8 with Atom processor (1.8GHZ) and 2GB system memory. Class A (31 students) participated in the study. No microphone or headphone was used in the study. Tuning in terms of levels of recording was done previously one by one. All the applications of the tablets used in the experiment worked well as a pilot experiment done by the author. The rate of recognition percentage (attempts succeeded/all attempts) was calculated by the students and we asked the students to report the rate after 30 min task with open-ended-feedback. The rate of recognition according to the location of the classroom was given in Figure 8 below, where ● represents the place where the participant sat down. The right picture shows the average recognition rate by each block.

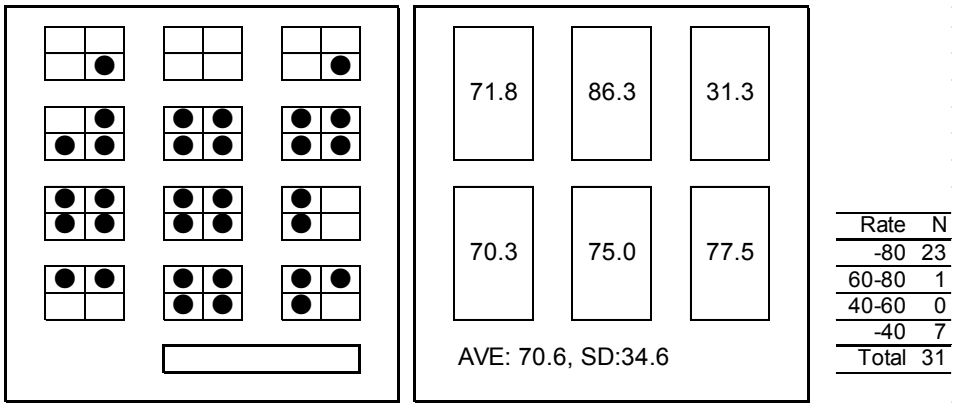



Figure 8 – Class A (Local Application on Tablet)

The result was lower than we had expected. It seems that class noise crucially affected the recognition of speech sound. Observing what was happening in the classroom, some tablets seemed

to be frozen or decoding process was interrupted by noise. Once it freezes, it did not work again how many times the students tried after that. When interviewed, the students with less than 40 percent answered that the task was really frustrating. Next, we arbitrary picked up 20 students and changed seat arrangements. We arranged so that one or two students could sit in an island. This, however, did not improve the situation. These results imply that it is difficult to integrate simultaneous tablet-based pronunciation training task in a classroom even if the class size is not so large.

### 3.3 Web application in CALL equipment

The second experiment involved web application with CALL equipment. Similarly, participants worked the same application installed in a Learning Management System (LMS); Glexa as indicated in Figure 3. Two different Classes (Classes B (N=40) & C (N=45)) participated in the study. They used a headphone in the experiment as indicated in Figure 4.

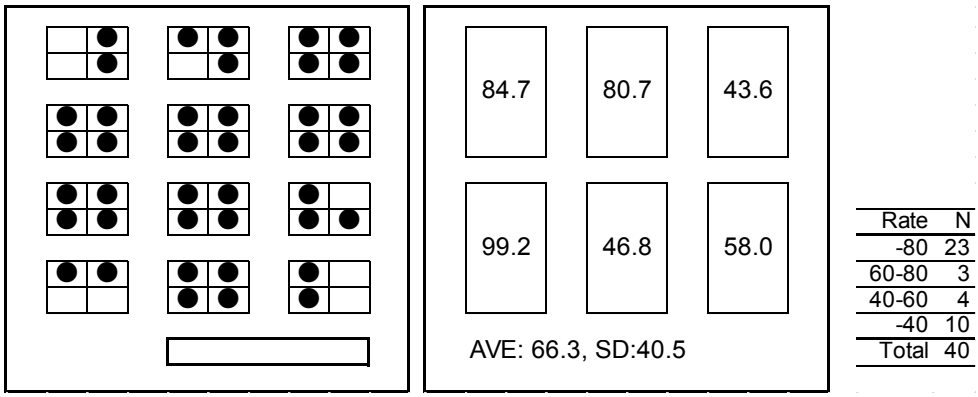



Figure 9 – Class B (Web Application in a CALL room)

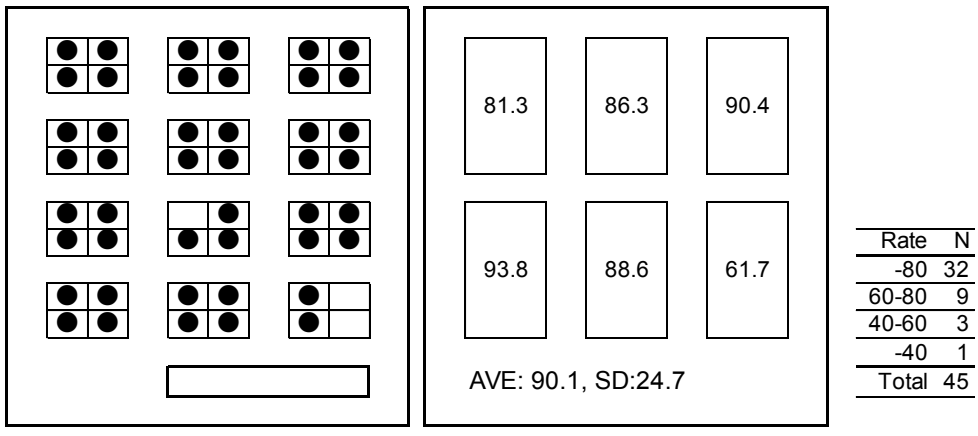



Figure 10 – Class C (Web Application in a CALL room)

The result is interesting in two ways. First, Class C, with 45 students, was more successful than Class B with fewer students. One reason might be the time of the class. Class B starts at 12:30 pm and Class C starts at 9:00 am. The network traffic might be lighter in the morning than in the afternoon, resulting in low recognition in the afternoon classroom. Secondly, in spite of the low accuracy in the case of Class B, the percentage has got better one month after, which imply the students got accustomed to the use of web application; volume of the speech, timing of articulation, location of microphone. More research is of course necessary to clarify the reason for that low recognition rate in the use of web application.

## 4. Conclusions

Considering the common use of tablets and effective incorporation of Voice recognition software,

noise is a great problem to overcome. This paper showed that more tuning or improvement is necessary for simultaneous use in the classroom. On the other hand, web application in a CALL room is a charming way for incorporation, because the software is directly installed into LMS server, enabling students to try outside the classroom. Our dream is the tablet-based software use not only in the classroom but also outside it. In his sense, the effective use of tablet software must be investigated in terms of hardware and instructional design.

## ACKNOWLEDGEMENTS

This study was partially supported by Grant-in-Aid for Scientific Research No. 26370656 and No. 24500176 from the Japan Society for the Promotion of Science (JSPS). We are grateful to an anonymous reviewer for his/her valuable comments given the manuscript.

## REFERENCES

1. Thomas, M., Reinders, H., & Warschauer, M. Contemporary computer-assisted language learning. London/New York: Bloomsbury; 2013.
2. Hiroyuki Obari, Hiroaki Kojima, Machi Okumura, Masahiro Yoshikawa, Shuichi Itahashi. Investigating the Effectiveness of Prontest Software to Train English Proficiency. SLATE2009-2009 ISCA International Workshop on Speech and Language Technology in Education; 2009
3. Ono, Y., Ishihara, M., & M, Yamashiro. Construction of wireless tablet-PC classroom for Computer-Assisted Language Learning in Japan. In Stephanidis, C., editor,. HCI International 2014 - Posters' Extended Abstracts: International Conference, HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014. Proceedings, Part II; 2014. p. 128-132.