

Effective Use of Text-to-Speech Technology for Japanese Learners of English

Kazuko AOKI

Research and Support Center on Higher Education for the Hearing and Visually Impaired,
Tsukuba University of Technology

Abstract: Synthetic speech has been used in various applications, but probably the most important and useful application of speech synthesis has been in reading and communication aids for the blind. The first commercial Text-to-Speech (TTS) application was the Kurzweil reading machine for the blind introduced by Raymond Kurzweil in the late 1970s. We found this system very useful for helping Japanese learners with visual impairments improve their English reading skills. Based on our success, we organized a small team to research the development of new English teaching methods for visually impaired students. After several trials, readKON, a reading support software equipped with English speech synthesis, was developed for use in our project. We also developed a computer-based vocabulary size test for the visually impaired named kobaTEST. In this report, we introduce these original software applications, and explain some of our findings. It is clear that these two TTS applications help sighted readers of English as well as the visually impaired.

Key Words: computer, English skills, Text-to-Speech (TTS), visually impaired, vocabulary size test

1. Introduction

1.1. What is TTS?

TTS refers to a Text-to-Speech conversion technology or engine that converts written text into a voice file. Speech synthesis technology has a rather long history. Hideo D. Harashima [1] notes, "According to Jacobson (n.d.) [2], the first text-to-speech system was completed at Bell Laboratories as early as 1968. Kilickaya [7] also states that the first TTS was implemented in the Speak and Spell handheld electronic learning aid by Texas Instruments in 1978." Until recent years, machine-made voices were of poor quality and were often ridiculed as "robot voices." Among teachers and researchers of EFL (English as a foreign language), synthetic speech has been evaluated poorly. However, Harashima has argued that corpus-based speech synthesis technology has developed so rapidly in recent years that it can now be considered a valid replacement for human voices for various purposes, including language learning.

1.2. TTS and Visual impairments

Synthetic speech has been used in various applications, but probably the most important and useful application field of speech synthesis is that of reading and communication aids for the blind. The first commercial application of Text-to-Speech technology (TTS) was the Kurzweil reading machine for the blind introduced by Raymond Kurzweil in the late 1970s. This machine consisted of an optical scanner and text-recognition software. This was

a significant technological development for the visually impaired because it provided users instant access to regular print materials. One of the system's most notable characteristics was the ability to change the speaking rate from 50 wpm (words per minute) to 500 wpm. At first, Kurzweil machines were mainly introduced to institutions for the visually impaired, including public libraries, because they were too expensive for individual users. However, modern systems are mostly software-based, so with a scanner and an OCR system, it is easy to construct a reading machine for any computer environment at a reasonable price. In addition, a TTS technology was developed to let blind users operate computers and read and write texts. This screen reader software is now applied to many languages, including Japanese.

1.3. Using Kurzweil machines as tools for English learning

About ten years ago, Aoki, a teacher of English at Tsukuba College of Technology (TCT), had several students who had lost their sight in their teenage years. Consequently, they could use neither printed materials nor Braille. Although they were learning the Japanese Braille system, they struggled to learn the English one because of lack of time. One student, a 20-year-old man called K, was strongly motivated to learn English, but he was irritated at the lack of learning materials. Face-to-face oral lessons with the teacher were his only opportunities to learn. Unfortunately, this time was limited, and it became necessary to find him a self-study environment. Then, we noticed that we had a Kurzweil reading machine in our classroom. Before this time, it had been somewhat neglected and used only for visitor demonstrations. Although the voice was a little robotic, it was clear enough to understand the English texts. Moreover, as it was designed for the blind, the student easily learned how to operate it. This was our starting point for using Kurzweil as a tool for English learning for the visually impaired. In the beginning, we focused on how to use the machine, because it was a sophisticated reading machine, not a tape recorder. Once the machine scanned text, the user could choose how the information was read. There were several options for reading speed and different reading modes, including by line, by sentence, and by word. Repeating, going back to the previous sentence, and reading through the paragraph were some of the other options.

For two years, we continuously developed new ideas in our weekly special English course. The newly blind student, K, practiced reading, shadowing, and dictation with the Kurzweil Reading Machine, in addition to learning English Braille. His reading speed in English Braille improved from 28.3 wpm to 54.3 wpm, while his Japanese Braille speed improved from 30.5 wpm to 35.3 wpm. (Aoki [3]) The support of the Kurzweil machine's synthetic speech facilitated his English Braille reading ability. In addition, his overall English ability improved dramatically. At the end of the two-year course, he passed STEP 2nd grade (Society for Testing English Proficiency, one of the most popular English proficiency tests held in Japan), which is the higher intermediate level.

2. Development of original reading support software

2.1. Why was readKON necessary?

We knew that the Kurzweil Reading Machine was useful for blind learners, but it was not easily applied to other types of students with visual impairments like low vision students. In response, we introduced the English version of a new screen reader called outSPOKEN at some of the personal computers in our classroom. This software was designed for blind people and severely visually impaired people. The screen reader enabled users to operate a PC as well as read and write English texts by support of synthetic speech and a refreshable

Braille display. This technology provided opportunities for reading improvement for students whose proficiency levels were rather high. However, many of our students were poor readers of English, and it was very difficult for them to use outSPOKEN because its reading speed was too fast. After several attempts of using new versions of Kurzweil and outSPOKEN, we realized that the problems stemmed from the fact that these tools were designed for native English speakers. They were too sophisticated for Japanese EFL learners with visual impairments whose proficiency levels were low. We needed a tailor-made reading support system. The original application software, readKON, was developed for this reason in 2002. Mr. Kunio Kondo, a teacher at Saitama School for the Blind, made this application software for our research group. ReadKON was given a good educational software award by the National Institute of Special Education, NISE. It can be downloaded from his homepage: <http://homepage2.nifty.com/k-kondo/readKON.htm>.

2.2. Features of readKON

Some of the main features of readKON are as follows:

- (1) It is a PC-based reading system developed specifically for ease of use by visually impaired students and their teachers.
- (2) The system utilizes Microsoft Speech SDK to assist students with reading fluency, reading comprehension, and vocabulary development.
- (3) This application allows the user to view a line (words, phrases, or a sentence, depending on the original text) on the computer screen while listening to the text being read aloud in English synthetic speech. In other words, it provides auditory and visual presentation.
- (4) The colors, fonts, and sizes of letters on the screen can be customized according to individual needs. The default setting is 100-point yellow letters on a black background.
- (5) It is geared to read one line of text in response to a reader's manual keyboard operation in order to allow him/her to control reading speed.
- (6) It is a keyboard-driven system, which means the reader does not need to use a mouse.
- (7) It reads any English reading material in a text file.

We used readKON in a variety of ways over several years, but it was not always welcomed by students. We knew that the biggest problem was the voice quality of Microsoft Speech SDK. Thus, in 2006 we introduced a new TTS engine produced by PENTAX, a Japanese company. The voice quality of this engine was good enough to satisfy our students. Since then, all the computers in our classroom have been equipped with this system, and students now use it for a variety of activities.

2.3. Reading skill training with readKON & some findings

As readKON is very simple software, it is applicable to various reading practices. The following programs helped slow readers improve their English reading skills.

1) Word-recognition test

Automatic word recognition is one of the most important factors of reading skill. According to Samuels (1979) [8] there are three steps in developing word-recognition skill. The first step is the non-accurate stage. The second stage is the accurate stage, and the third stage is the automatic stage. Students at the first stage have difficulty recognizing words and cannot read them correctly. Those at the second stage can read words correctly but slowly. They often read sentences word by word. The last group of students, those who have reached the third stage, read words correctly and instantly at their speaking speed or faster.

We developed a computer-based word recognition test for the visually impaired using readKON. In the test, the TTS voice was not used and students were told to read the five groups of words as fast as possible. The groups consisted of 30 words of different lengths, ranging from 3-letter words to 7-letter words. The number of readable letters per second (LPS) in each word group was calculated. Analyses of the test results for visually impaired students revealed some interesting patterns. Considering the recognition rate of good readers compared with that of our students, there are three types of readers. (Figure 1)

There were two types of slow readers among our students. Although the word-recognition speeds of some students (LV1) were nearly half that of a good reader with normal vision, their word-recognition rates in different groups (3-letter words to 7-letter words) were quite similar to those of good readers. This means that their levels of automatic word recognition are rather high, and in fact, they are slow readers but not poor readers. The other group of students (LV2) is considered more typical. The data show their word-recognition levels are quite low and they read words letter by letter. Thus, this word-recognition test indicates learners' levels of automatic word recognition, which is a basic English ability.

2) Word-recognition training

Although speech synthesis is not used in the word-recognition test, it is very useful for word-recognition training. A teacher can prepare any type of word list for students. Speech synthesis allows students to see the words on the screen and hear how they sound. In this way, they can learn pronunciation and repeat the words as many times as they want. They can also improve their word-recognition skill with practice. The following is one of the case studies conducted in our project.

3) Case study: A struggling reader

M is a 29-year-old male student who entered the acupuncture program to get a national license for Acupuncturists of Japan in 2004. He finished the regular high school course at the age of 18 and had worked for several years. Therefore, he had been out of school for almost 10 years. In addition, he had had a bitter experience learning English previously.

M avoided taking the proficiency test when he was a first-year student and frequently missed English classes. It was difficult to measure his exact level of English proficiency; however, it might have been as low as that of a novice. The following are the results of his first word-recognition test. Figure 2 shows changes in word-recognition rate (LPS) and Figure 3 shows M's accuracy rate, which refers to how many words he read correctly.

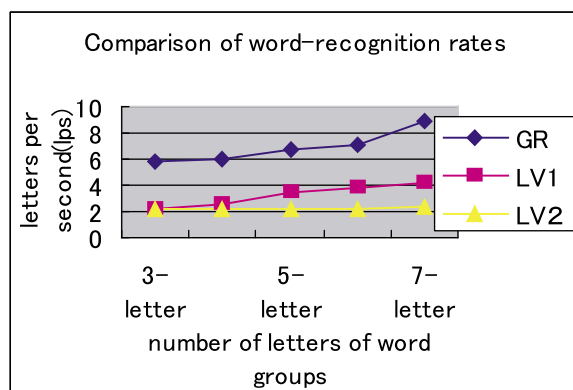


Fig. 1 Comparison of word recognition rates
 GR: sighted university teacher (good reader)
 LV1: low vision student (intermediate)
 LV2: low vision student (poor reader)

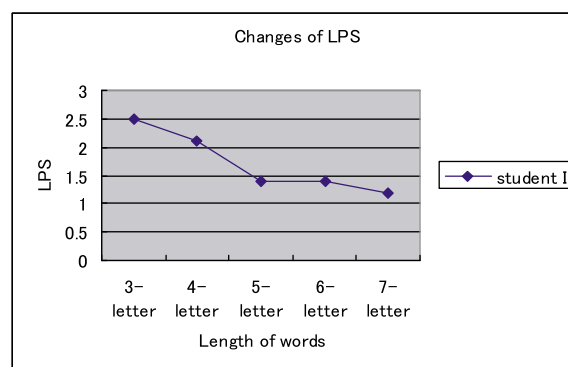


Fig. 2 Change in LPS of student M

The two figures show that M is a typical poor reader. The next figure (Figure 4) shows the difference between his LPS-pattern and that of a good reader.

The results of this test made it clear that he should start word-recognition training. Several word lists were prepared for these training sessions. M practiced them on a PC using readKON for an hour each session. At the end of the session, the same word recognition test was performed. This training took place over five sessions. Changes in LPS and accuracy rates are shown in Figures 5 and 6, respectively.

After comparing the first and fifth trials, it is clear that M's recognition speed has improved. However, he still has difficulty recognizing 6-letter and 7-letter words.

On the other hand, M's accuracy rates steadily improved and he was able to correctly read most of the words and phrases given in the training sessions.

These results show that this type of word-recognition training is useful for poor readers like M. We also observed that his concentration level gradually improved during training, and he looked comfortable and confident. However, in terms of automatic word-recognition skill, M was still at a low level. It was clear that he needed more time for training.

4) Speech practice with TTS

English oral ability is a valuable assessment tool for general English classes because the process of oral presentation includes writing, reading, listening, and speaking. For the

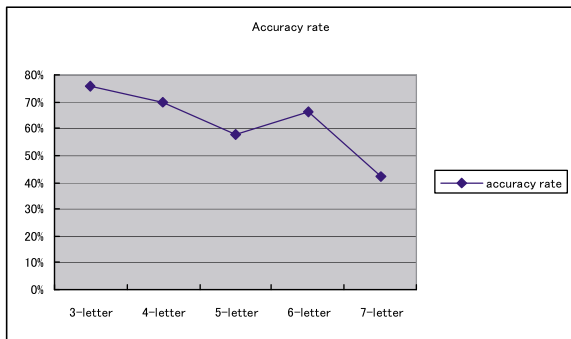


Fig. 3 Accuracy rate of student M

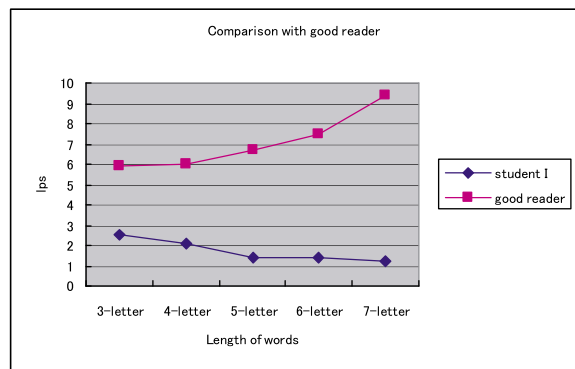


Fig. 4 LPS comparison between student M and a good reader

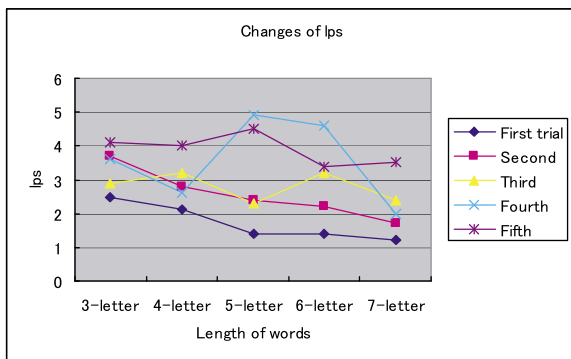


Fig. 5 Changes in M's LPS after training sessions

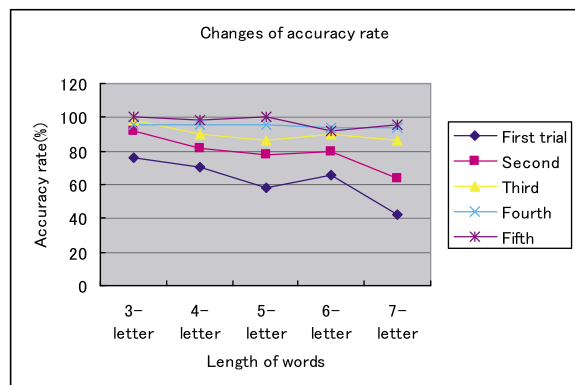


Fig. 6 Changes in M's accuracy rates after training sessions

last several years, the writer gave almost the same tasks to second-year students with visual impairments. The following are details of the task given to students.

1. The title of the speech is "My Hometown."
2. The script should consist of more than three paragraphs.
3. The length of the script should be around 500 words.
4. The drafts of the script should be submitted to and checked by the teacher at least three times.
5. The speech should be recorded on an audio tape and delivered at the final class.

Considering the purpose of this report, we will concentrate on the problems of speech delivery encountered by our students, i.e. Japanese EFL learners with visual impairments.

Accuracy, fluency, and speed were set as the evaluation items for the students' speech deliveries. There were several distinct characteristics of the students' recorded speeches. The first characteristic was slow speech, which included frequent pauses and lack of English rhythm. Many students took more than 10 minutes to finish their speeches, which is less than 50 wpm. The second problem was incorrect word pronunciation, which often hindered listeners. These problems inevitably point back to lack of fluency as a whole. One of the biggest causes of these problems was insufficient practice time, as students usually spent nearly 80% of their time completing the final draft. We realized that these students needed much more time for reading practice than sighted students, because their reading materials were printed in Braille or large print.

In 2006, the program was revised to improve the situation. We decided to introduce readKON for practice. It was clear that our students needed some kind of speech model, including bottom-up training of word pronunciation and improvement of fluency. The reading support software, readKON, offered speech models with TTS at any time. Additionally, we set a longer period for practice and assigned students a target time of five minutes for their recorded speeches.

(Results)

Various writing tasks were given from the start of the course, including paragraph writing, 50-word impromptu essays, and 100-word structured essays. The students were given oral presentation tasks in the second term. The procedure was almost the same as the one given in the third term mentioned above, though the title was "A Great Person" and the length of

Table 1: Results of evaluation and speech speed

St	Visual impairment	Second Term		Third Term		Reading Speed 2005 WPM	Speech Speed 2007 WPM
		A	F	A	F		
A	totally blind	3	3	4	4	50	78
B	totally blind	3	2	4	4	37	85
C	light perception	1	1	2	2	27	75
D	low vision	4	4	5	5	60	141
E	low vision	2	2	4	3	56	88
F	low vision	3	2	4	4	45	101
G	low vision	2	1	3	3	62	78
H	low vision	1	1	2	2	30	71
I	low vision	2	1	3	2	26	47

Notes:

A: Accuracy grade (Excellent 5-4-3-2-1 Poor)

F: Fluency grade (Excellent 5-4-3-2-1 Poor)

WPM: words per minute

the draft was 300 words. However, the TTS system was not used for reading practice and many of the students struggled to finish their recordings. Data on the evaluation items of the two recorded speeches, "A Great Person" and "My Hometown," were compared for each student. Since the data were too small and the students' levels of English proficiency and visual impairments were too varied for any statistical analysis, the raw data are shown in the following table.

Every student made some degree of progress in terms of accuracy and fluency. The next figure shows their progress.

We should consider student C in more detail. Student C was a non-Braille reader who usually used a PC supported by a Japanese TTS system for his study. He could read and write Japanese texts very well, but his PC system was not equipped with an English speech engine. Therefore, he usually read and wrote English texts with the Japanese TTS engine. As a result, he was not well accustomed to authentic English pronunciation. He tried to practice his oral presentation with readKON but he gave up quickly because he could not understand the English speech generated by readKON. He decided to use the Japanese TTS in his PC, and as a result, his performance in his recorded speech was not improved at all compared to the previous speech.

The next interesting finding was the change in all students' speech speed. Each student finished his/her speech within five minutes, even though we expected it would be too difficult for some of the students. However, comparison of speech speed between the two recorded speeches was difficult, because in some cases the quality of the first speech done in the second term was too poor to be evaluated due to inaccurate pronunciation, frequent pauses, and an unclear, small voice. As a result, we evaluated data from the students' reading speed taken two years previously. This showed that all the students were initially slow readers with reading rates around 50 wpm or less. Figure 8 shows the reading speed in 2005 and the speech speed in 2007 for each student. Five of the nine students doubled their speech rates (wpm). The others also increased their speed, improving more than they had expected.

(Discussion)

There were very interesting differences between the two recorded speeches. Every student made some progress in three evaluation fields, which indicates that the TTS system effectively supported students' performance. Of course, it was very important regarding effective use of

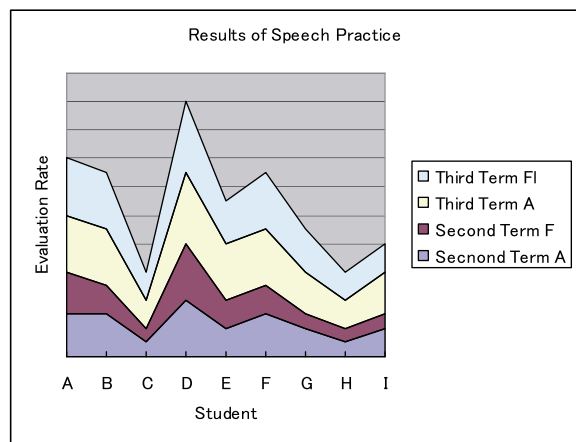


Fig. 7 Results of evaluation rates

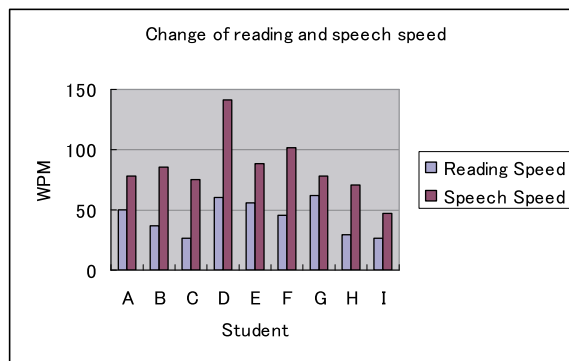


Fig. 8 Changes in reading & speech speed

TTS system but it was also expected, as it would be. More interestingly, we found the quality of students' voices improved in the second speech. In the first speech, their voices were low and unclear, but in the second speech they sounded clear and powerful. It was apparent that the TTS practice decreased their anxiety about speaking or reading in English.

3. Development of Vocabulary Size Test (kobaTEST)

3.1. Assessing vocabulary sizes

It is important to assess students' proficiency levels at the beginning of the course. However, the results of such a test are sometimes questionable because they seem to be lower than expected. In most cases, this problem stems from students' visual impairments. Although the time is extended and test materials are prepared in ordinary print, large print, or Braille, depending on their needs, it is difficult for them to complete the test within the time set by the examination administrator. We thought that a simpler and less stressful test was needed to assess students' ability. A vocabulary size test is one such option. Several vocabulary size tests have been developed for Japanese EFL learners. However, we found that it was difficult for low vision students and blind students to complete standard vocabulary size tests because they usually need a lot of time and stamina.

Therefore, we developed a computer-based vocabulary size test for the visually impaired. This PC-based voc-size test was named kobaTEST. (This application was made by Mr. Makoto Kobayashi, one of our project members.)

3.2 Features of kobaTEST

KobaTEST was designed with the same concepts used in readKON, which provided auditory and visual presentation.

Twenty words are selected randomly from a word list in each session. A target word appears on the display along with the voice of the English speech synthesizer. At the bottom of the screen are three boxes that say, START, I KNOW, and I DON'T KNOW in Japanese (Figure 9). Students are expected to indicate whether they know the meaning of the target word by selecting one of the boxes with the mouse or keyboard. The test is simple enough to finish within one minute even for low vision students. Information for each test and the students' responses are recorded in the computer.

Five different levels of word lists were prepared for the test: Level 1 (the lowest level), Level 2, Level 3, Level 4, and Level 5. These lists were based on the JACET List of 8000 Basic Words (2003, JACET), which was specifically designed for Japanese college EFL



Fig. 9 Display of kobaTEST

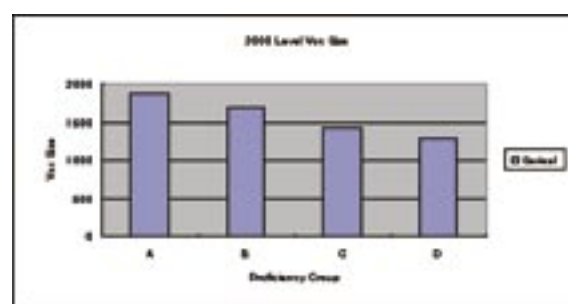


Fig. 10 Vocabulary sizes of 2000-Level

learners. Each word list consists of 1,000 words.

3.3. *What we found—the importance of 2,000 basic words*

Data collection on the vocabulary sizes of our students using kobaTEST began in 2002, and the research is still ongoing. So far, the results show that visually impaired students with very small vocabulary sizes between 1,000 and 2,000 words are inevitably slow and poor readers. Figure 10 shows some of the data collected in 2006 and 2007. The number of students was 54. We divided them into four groups according to their English proficiency levels; group A was the highest level group and group D was the lowest. Their vocabulary sizes exactly match their proficiency levels. This means that a student's proficiency level can be predicted by looking at the results of kobaTEST.

3.4. *Applications*

KobaTEST can also be applied to a variety of vocabulary-building practice lessons by inputting an Excel-based word list into kobaTEST. Each time, kobaTEST shows 20 randomly-selected words with voice support. Learners can use this system for practicing pronunciation and spelling. The strongest point of kobaTEST is that the test words change each time. In general, printed vocabulary tests have only one version, and test words are set in the same order. Thus, if one tries to assess student progress in a rather short period, the test cannot be used because students might remember the previous test. In our class, students can choose an appropriate level word list (Level 1–5) and learn it independently using kobaTEST. They try to take the test time to time when they want to know their progress.

4. Conclusion

In this report, we looked at the history of our projects, which aimed at developing new teaching methods for Japanese college students with visual impairments. We focused on TTS systems applied to various teaching activities. In the early stages, we tried to utilize commercial TTS systems for our students. Through pilot studies as well as real classroom situations, we realized that we had to develop original application software. ReadKON and kobaTEST were the two main systems developed for our use. Our research is now geared toward developing appropriate programs for these systems.

Many valuable findings were extracted from real classroom situations in which teachers and learners struggled. In fact, TTS technology has given us a completely new learning environment for EFL learners with visual impairments. Considering the dramatic attitude changes in our students, especially poor readers of English, we feel certain that this kind of system should be useful to any student who is struggling with language learning. We hope our small study will provide useful insights and strategies for other language teachers.

Acknowledgements

Special thanks are due to the following members of our project:

KATO Hiroshi, Professor of Research and Support Center on Higher Education for the Hearing and Visually Impaired, Tsukuba University of Technology

KOBAYASHI Makoto, Associate professor of the Department of Health, Tsukuba University of Technology

KONDO Kunio, Teacher at Saitama School for the Blind

Notes

1. STEP (Society for Testing English Proficiency): a large scale English proficiency test

conducted throughout Japan, mainly in high schools and other educational institutions. It offers special opportunities for disabled learners, including blind people.

<http://stepeiken.org/>

2. Kurzweil Technology

<http://www.kurzweiltech.com/kesi.html>

3. Speech Synthesis voiceTEXT by PENTAX

<http://voice.pentax.co.jp/>

4. Screen Reader Software: xpNAVO

(This newly developed screen reader software utilizes the PENTAX engine.)

<http://www.knowlec.com/product/navo-catalog.html>

5. JACET (2003), JACET List of 8000 Basic words, The Japan Association of College English Teachers

References

- [1] Hideo D. Harashima, *Electronic Journal of Foreign Language Teaching* , 2006 Vol. 3, No. 1, pp. 131-135
- [2] Jacobson, K. (n.d.). Approaches to speech synthesis. Retrieved May 19, 2006, from <http://umsis.miami.edu/~kjacobso/speechsynth/speechsynth.htm>
- [3] Kazuko AOKI, Improving a newly blind student's English reading skills by Kurzweil Reading Machine, *Tsukuba College of Technology Techno Report*, 1999, Vol.6 167-173
- [4] Kazuko AOKI, Are Visually Impaired Students Slow Readers? - What reading support software can do for them?, *Eurocall 2003*
- [5] Kazuko AOKI, "Developing English Reading Support System and Vocabulary Size Test for Japanese Visually Impaired Students - What Computers can help them to study English? - ", 12th ICEVI World Conference 2006, Abstract Reference Number EA 028
- [6] Kazuko AOKI, Hiroshi Katoh, Makoto Kobayashi, "Effective Use of Text-to-Speech Technology (TTS) for EFL learners of Japan," *EUROCALL 2007*, p.66
- [7] Kilickaya, Ferit. 'Text-to-speech technology' : What does it offer to foreign language learners? 2006 *CALLEJ Online*, 7(2). Retrieved January 31, 2006, from <http://www.tell.is.ritsumei.ac.jp/callejonline/journal/72/Kilickaya.html>
- [8] Samuels, S.J., The methods of repeated readings, *Reading Teacher*, 1979, 32, 403-408