

A Comparative Study on Reading Aloud: Instructions by Text-To-Speech Synthesis Sounds and a High School Japanese English Teacher

Harumi Kataoka
(Graduate School, Kansai University)

Makiko Ito
(Osaka Prefectural Kashiwara-higashi High School)

1. INTRODUCTION

The curriculum guidelines (2010) of the Ministry of Education, Culture, Sports, Science and Technology (MEXT) required Information and Communication Technology (ICT) into practical use in not only Information lessons, but also each subject lessons. For example, the curriculum guidelines recommend using audio CDs, DVDs, and computers during foreign language learning in Japanese elementary school education.

With the advancement of computer software, various styles in foreign language teaching have been developed since the early 2000s, including Text-To-Speech (TTS) synthesis technology. TTS refers to digital speech synthesis technology using audio database of native English speakers. Users can produce TTS synthesized speech sounds automatically by typing English words on their computer keyboards. By using TTS, teachers can control English speech sounds of (a) the speed, (b) the pause, (c) the duration, and (d) even choosing TTS speakers. Then they can produce audio English language learning materials to fit their students' English proficiency (Kataoka, 2009).

However, there are not enough research results on how to use TTS speech sounds in the classroom, especially for Japanese high school students. This may be the reason why some native and non-native English teachers in Japan have negative perceptions about using TTS synthesized speech sounds during their classroom lessons.

In short, more studies on the use of TTS speech sounds as English language learning materials are needed in Japanese English as a Foreign Language (EFL) education.

2. LITERATURE REVIEW

2.1 English Education for EFL Learners

Palmer (1921), a British linguist, who endeavored to build the foundation of English language education in Japan, insisted that students needed to develop their listening abilities to understand English speech (pp. 400–401). In order to do so, Hatori (1977, p.74) recommended that Japanese students need to understand both English spelling and English pronunciation of each English word on an individual basis. He also pointed out that using English speech sounds during English lessons would enhance Japanese students' learning motivation. Widdowson (1978, p.57) suggested that four skills—speaking, listening (or

understanding speech), reading, and writing—are closely related to each other in developing English acquisition among learners. Kohno (1984) carried out studies which proved that English speech sounds helped Japanese students understand English. He mentioned that the Japanese and English languages present differences in their prospective linguistic structures and pronunciations, and suggested that using English speech sounds as much as possible during English lessons is of utmost importance. These previous studies have indicated that audio information and visual information are closely related to understand English.

2.2 Reading Aloud for Japanese EFL learners

Reading aloud, which is called “*Ondoku*” in Japanese, is one form of English learning activities. As suggested by Widdowson concerning four skills, a reading aloud activity would be practical English learning strategy. There have been some studies on reading aloud in Japanese EFL education.

Watanabe (1988, 2009) conducted a series of comparative studies about reading aloud. In Watanabe (2009), the participants were categorized into two groups: (a) a reading aloud group consisting of 44 second-year Japanese high school students, and (b) a non-reading aloud group consisting of 45 second-year Japanese high school students. The results showed that students in the reading aloud group obtained higher test scores and could read and understand English sentences faster than the learners in the non-reading aloud group.

Suzuki (1998) conducted a comparative study on reading English sentences aloud during classroom lessons in a high school. The results of the English Test of National Center Test for University Admissions revealed that students who had participated in reading aloud showed higher test scores than those without reading aloud during their English lessons. His results suggest that reading aloud during classroom lessons helped promote Japanese high school students’ English proficiency.

Ishihara (2002) also conducted comparative research on reading aloud. The participants were 80 second-year high school students who were divided into two groups. The results of the cloze test revealed that 40 students in a reading aloud group marked higher scores than 40 students in non-reading aloud group. The results show that the activity of reading aloud is effective in keeping the memory of the text which they read.

Ikeda and Takeuchi (2003) examined a reading aloud activity conducted by 37 Japanese university students. They read aloud two kinds of English passages, and were then asked to answer a cloze test. The findings in this study showed that their reading aloud abilities were correlated with their English proficiency levels. This suggested that reading short and simple English sentences aloud would be appropriate for Japanese EFL learners.

Miyasako (2002) investigated a reading aloud activity for 40 high school students during their English lessons. The results showed that reading aloud improved students’ English grammar, vocabulary, listening and reading abilities, which implied that this activity. Reading aloud could help develop high school students’ English proficiency. Miyasako (2007)

also concluded that reading aloud is an effective English learning method for learners to acquire comprehensive English proficiency.

Kitamura (2007) researched the effect of reading aloud while writing English sentences repeatedly. A total of 158 Japanese high school students—36 first graders, 59 second graders, and 63 third graders—participated in the treatment groups. Their regular school tests and two kinds of tests offered by The Society for Testing English Proficiency (STEP), Inc. were used to monitor the progress of the participants' English proficiency. The results revealed that test scores were positively correlated with the time spent in reading aloud. These previous studies suggest that reading aloud is an effective learning strategy for Japanese EFL learners to acquire English. Reading aloud activities appear to be effective for Japanese high school students to learn and memorize English grammatical structures.

2.3 TTS Studies in EFL Education

There have been some research on TTS synthesis technology in EFL education conducted in Japan and overseas. Jones, Berry, and Stevens (2007) conducted an experiment using a TTS voice. The results showed that both native and non-native speakers of English did not perceive any differences between TTS speech sounds and human speech sounds. The findings indicated TTS's usability for English education in Australia. Chew (2009) reported that a written news text was digitally converted into TTS speech sounds, and used as an English language test material in Singapore. The finding suggested that the TTS system is useful for language teachers who want to create audio language learning materials by themselves.

In Japan, Azuma (2008) described how the synthesized English speech generated by TTS technology could be used, and discussed its introduction in an EFL class in a Japanese university. He also pointed out that TTS speech sounds differ depending on which TTS system is used. A study conducted by Yoshida (2008) on an electronic phrasal verb workbook paired GIF animation together with TTS synthetic speech. The result showed that the moving images combined with TTS speech sounds were effective in learning English among Japanese university students. Kataoka (2009) investigated whether Japanese undergraduate students and university graduates perceived any differences among three types of speech sounds: (a) TTS speech sounds, (b) the voice of a native British speaker, and (c) an American speaker. The results concluded that participants did not find any differences among the three types of speech sounds. The results indicated that TTS speech sounds have sufficiently high quality to be used as audio materials for Japanese EFL learners.

Hirai and O'ki (2011) examined the quality of TTS synthetic speech by comparing it with natural human speech. The participants were Japanese university students who were divided into groups. The results of listening tests showed that TTS speech sounds were perceived to be almost as natural as human speech by both upper and lower English proficiency groups, and more students in the lower level group tended to prefer TTS speech sounds.

3. COMPARISON BETWEEN TWO KINDS OF INSTRUCTIONS

3.1 Purpose

The results of previous TTS studies (Azuma, 2008; Chew, 2009; Hirai and O'ki, 2011; Jones, Berry, and Stevens, 2007; Kataoka, 2009; Yoshida, 2008) show that there are not so much difference in voice qualities between TTS speech sounds and English speech sounds by human speakers. Because TTS speech sounds are originally created using native English speakers' voices. TTS speech sounds appear to be useful for both Japanese EFL learners and English language teachers who really want to create audio language learning materials that fit their students' English proficiency level. Unfortunately, some native and Japanese English teachers have negative perceptions about using TTS speech sounds during their classroom lessons, especially high school education in Japan. Most of the English teachers in Japan are not aware of the usability of the TTS systems. Moreover, there are not enough research results to show how TTS can be used in the classroom, especially for Japanese high school students. Therefore, there is a need for more experimental studies of TTS speech sounds to be conducted for Japanese high school EFL education.

This study investigated whether TTS speech sounds could be used as an English language learning material for Japanese high school students to learn basic English verbs through reading aloud activities.

With the purpose of revealing TTS speech sounds as an audio English language learning material, TTS speech sounds were recorded and produced on TTS CDs¹ to be used in the present study. The research question (RQ) is as follows.

RQ: Is there any significant difference in basic English verb learning between the two kinds of reading aloud instructions: (a) a TTS CD style instruction, and (b) a verbal instruction made by a Japanese English teacher?

If there is no significant difference between the two kinds of instructions in their effectiveness, the TTS speech sounds in this study would prove to be of high quality and usability as an audio language learning material for Japanese EFL education.

3.2 Review of Ten Basic Verb Words

In order to have the students review basic verb words, which they had learned in junior high schools, following ten basic English sentences cited from Otsuka's² online English language learning materials (2009) were used in this study.

1. Are you a student?
2. Are you from Sapporo?
3. Is this a book?
4. Is Tokyo Disneyland in Tokyo?
5. Is there a pen on the desk?
6. Is Japanese spoken in China?

7. Are there any good restaurants in your city?
8. Do you have any good restaurants in your city?
9. Do you like English?
10. Do you study English every day?

3.3 TTS System in This Study

The TTS systems are composed of two items: (a) a TTS engine, and (b) a TTS audio database (Dutoit, 1997). The free TTS software application Balabolka Version 1.9.0.241, which can be downloaded free of charge, was used as a TTS engine in this study. With Balabolka, it is possible to define the duration of a short pause to within 1/1000 of a second. As a TTS audio database, “Mike & Crystal”, who were male and female names with synthesized American English voices respectively, and produced by AT & T, was also installed in a personal computer. These two items were used to produce a TTS system for this study (Figure 1).

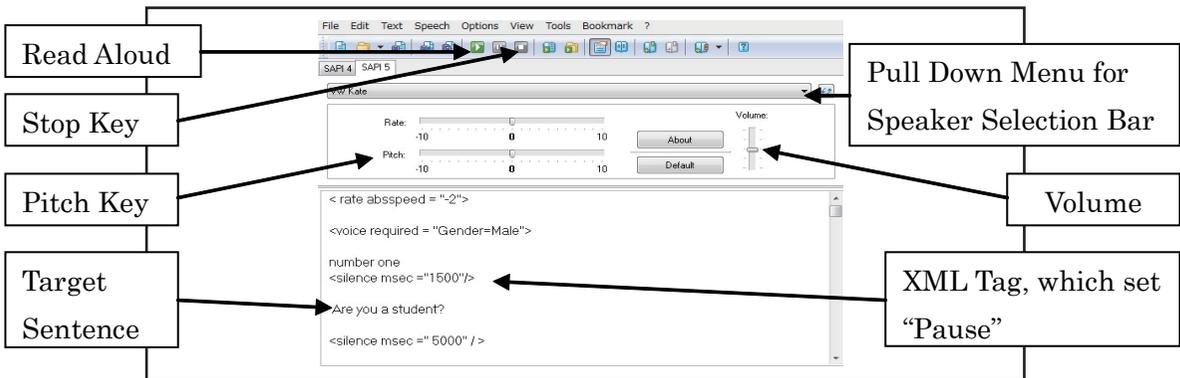


Figure 1. A screen of the TTS system.

4. Experiment 1: Speech Analyses

In order to examine the speech quality, two types of speech analyses were conducted: (1) acoustic analyses by a speech analyzer “Praat”, and (2) speech evaluation by raters.

4.1 Results of Acoustic Analyses by a Speech Analyzer “Praat”

Three kinds of speech sound (1) male TTS speech sounds, “Mike”, (2) female TTS speech sounds, “Crystal”, and (3) speech produced by a Japanese female English teacher were used in this study.

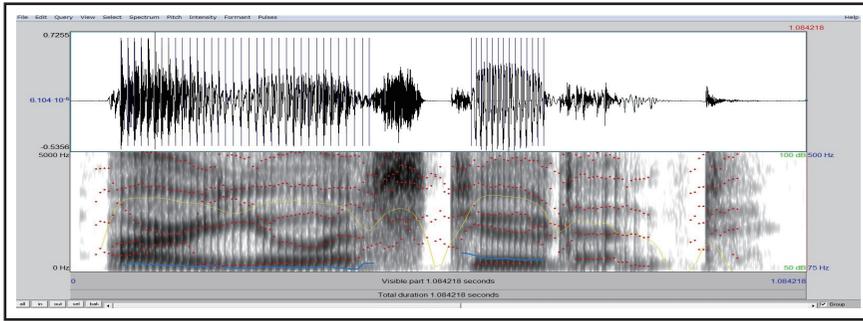


Figure 2. A screen image of the speech analyzer “Praat”.

We analyzed duration, the maximum pitch, the minimum pitch, and pitch range to identify features of speech sounds. Table 1 shows average rate of three kinds of speech sounds in this study.

Table 1. The Average Rates of Ten English Sentences by Three Kinds of Speech Sounds

Speech sounds	Duration	Pitch (Max.)	Pitch (Min.)	Pitch Range
TTS Male	1.72 ms	222.37 Hz	79.71 Hz	142.66 Hz
TTS Female	1.98 ms	239.60 Hz	125.30 Hz	114.30 Hz
JET	2.01 ms	372.35 Hz	165.95 Hz	206.40 Hz

Note. A Japanese female English teacher’s speech sounds shows “JET”.

As Table 1 indicates a Japanese female English teacher used higher pitch than TTS male/female speech did. The duration of speech of the Japanese female English teacher was the longest among three kinds speech sounds, which indicates that her speech rate was the slowest among three speech materials.

4.2 Results of Speech Analyses by Raters

Seven English teachers participated in this study as raters and they were asked to judge three kinds of speech sounds. The raters consisted of six Japanese English teachers and one assistant language teacher (ALT). They worked at a Japanese high school where the experiment of this study was conducted. They were asked to listen to the three kinds of speech sounds, and then they evaluated three speech samples based on its intelligibility, comprehensibility, and naturalness as spoken English using the five-point Likert scale style (Takeuchi, 2003, p. 254). Table 2 shows the results judged by seven raters in this study. The “ α ” in table 2 means “Cronbach’s coefficient alpha”. Generally speaking, it is said that if α value shows over 0.70³, there is enough “Inter Rater Reliability”. The α values of three kinds of speech sounds in this study showed over 0.90, and revealed there were no significant difference between seven raters’ scores.

Table 2. The Results of Three Kinds of Speech Sounds by Seven Raters

Speech sounds	<i>M</i>	<i>SD</i>	<i>α</i>
TTS Male	3.12	0.92	0.93
TTS Female	3.38	0.88	0.90
JET	3.35	0.99	0.96

Note. The “*α*” means “Cronbach’s coefficient alpha”, which shows “Inter Rater Reliability”.

Table 2 also showed that TTS female speech sounds gained the highest mean score of three speech sounds. This result suggests that TTS female speech sounds had higher speech quality than a Japanese English teacher’s speech sounds, and thus easy to understand for seven raters.

5. Experiment 2: Basic Verb Words Learning by Reading Aloud

It became evident from the results of experiment 1 that three kinds of speech sounds in this study had enough speech quality. These speech sounds were used as listening materials for reading aloud activities in this study. The results of previous studies on reading aloud activities (Ikeda and Takeuchi, 2003; Ishihara, 2002; Kitamura, 2007; Miyasako, 2002, 2007; Suzuki, 1998; Watanabe, 1988, 2009) show that this activity is one of the effective English learning methods for Japanese EFL learners. This is the reason why reading aloud was adopted as a learning method in the current study. Results of reading aloud were measured with the cloze test styles, as were used by most of previous reading aloud studies.

5.1 Participants

This study was conducted in January 2010, and a total of 151 male and female Japanese high school students participated in this experiment. The participants were all first year students from four classes in the same public high school in the Kansai area. They were divided into two groups according to their regular classes. Group A consisted of 59 students (25 boys and 34 girls), and Group B had 56 students (24 boys and 32 girls). All of them were born in Japan and had been studying English for more than three years. None of them had stayed in an English-speaking country or had any experience of participating in this kind of experiment. We ensured that the participants remained anonymous throughout the study.

When all participants took their high school entrance examinations, the average rating points of school evaluation, *Naishinten*, was 4.2 points on a zero-to-ten scale. Therefore, it is speculated that most of the students in this high school had not mastered basic English grammar, which was taught in junior high schools, they had to spend a few minutes reviewing junior high school level English grammar during their senior high school English lessons.

5.2 Research Design and Procedure

To investigate the effectiveness of TTS speech sounds as an audio English language learning material for Japanese high school students to study basic English grammar, the following four phases of the learning process (Table 3) were adopted in this research. The classroom practice consisted of two reading aloud tasks for six lessons. In the reading aloud task, students read English sentences silently, while listening to English speech sounds. After that, students repeated exactly what they heard, while reading the English sentences. Table 3 shows the timeline for this study.

Table 3. Timeline for This Study

Phase	Day	Activities	English Sentences	Audio Instruction	
				Group A (<i>n</i> = 59)	Group B (<i>n</i> = 56)
1	1	Pre 10-item Cloze Test	No. 1 – No. 10	No Audio Instruction	No Audio Instruction
2	2	Reading Aloud (1)	No. 1 – No. 5	TTS CD Sounds	A JET's Voice
	3	Reading Aloud (1)	No. 1 – No. 5	TTS CD Sounds	A JET's Voice
	4	Reading Aloud (1)	No. 1 – No. 5	TTS CD Sounds	A JET's Voice
3	5	Reading Aloud (2)	No. 6 – No. 10	A JET's Voice	TTS CD Sounds
	6	Reading Aloud (2)	No. 6 – No. 10	A JET's Voice	TTS CD Sounds
	7	Reading Aloud (2)	No. 6 – No. 10	A JET's Voice	TTS CD Sounds
4	8	Post 10-item Cloze Test	No. 1 – No. 10	No Audio Instruction	No Audio Instruction

5.3 Materials

5.3.1 Cloze Test

English sentences in the cloze test were selected from English learning materials published on the Web page by Otsuka² (2009). Participants in this study were unable to master basic English grammar taught in junior high school level. We decided to use all English sentences offered by Mr. Otsuka on his Web page, and his permission was obtained to use them in this study. Our aim is for the participants to be able to understand basic English grammar through reading aloud during classroom practice in this study.

The cloze test styles were used by previous studies on reading aloud activities (Ikeda and Takeuchi, 2003; Ishihara, 2002; Kitamura, 2007; Miyasako, 2002, 2007; Suzuki, 1998; Watanabe, 1988, 2009) to check the progress in learners' English proficiency. In this study, the same 10-item cloze tests at pre- and post-test were used to check the learning process of high school students (Appendix A). The cloze test in this study was a multiple-choice question format. Students were asked to select the best answer among four choices for each blank, and wrote one answer for each question on their answer sheets. The maximum number of points in the cloze test was ten. This type of cloze test with four multiple choice questions is used in the university entrance examination as (1) the first-stage English Test of the National Center Test for University Admissions and (2) written tests in many private universities. We consider that high school students in this study should be offered this type of test as much as possible to have learners get accustomed to this pattern of university entrance examination tests. This

is why the cloze test with four multiple-choice questions was used in this study.

5.3.2 Materials for Reading Aloud Task: Activity Sheets and TTS CDs

In order to compare the effectiveness of two kinds of instructions, (a) TTS CD style instruction and (b) instructions using only a Japanese English teacher's voice, were carried out with the two groups.

All students in Groups A and B were given either TTS speech sounds or a female Japanese-English teacher's voice for each English sentence. During the reading aloud practice, students read English silently while listening to English speech sounds, and then they repeated English sentences exactly as they had heard. The pauses between the English sentences were set 5,000 ms (= five seconds) so that they could repeat each English sentence together in the lessons (Figure 1).

The students in Groups A and B had to learn ten English sentences—1 through 10—in the cloze test (Appendix A) by reading them aloud. It appeared difficult for the students in the two groups to learn ten English sentences in one lesson. Therefore, five of the ten English sentences in the cloze test were learned in three lessons, each under Reading Aloud Tasks (1) and (2) respectively (Appendix B).

According to previous studies (Hatori, 1977; Kohno, 1984; Palmer, 1921), knowing both pronunciation and meaning of each English and Japanese word would facilitate Japanese high school students to understand English. Each group was given "Activity Sheets (1) and (2)" (Appendix B), which were used for the Reading Aloud Tasks (1) and (2) respectively.

In each activity sheet, five English sentences in the cloze test were printed in both English and Japanese. This way, students could easily understand English sentences by aid of the Japanese sentences. In the cloze test, there were ten English sentences that were all question sentences. In the reading aloud task, students learned how to answer each question by using activity sheets and reading aloud practice.

The following is a description of the two-step learning process adopted for the reading aloud task:

1. The first step was a "reading and listening" phase. Students were given an "Activity Sheet". They read this sheet silently while listening to either TTS CD sounds or a JET's voice.
2. The second step was a "repeating" phase. After the first step, all students repeated in unison each English sentence exactly what they had heard, while reading each English sentence on the activity sheet.

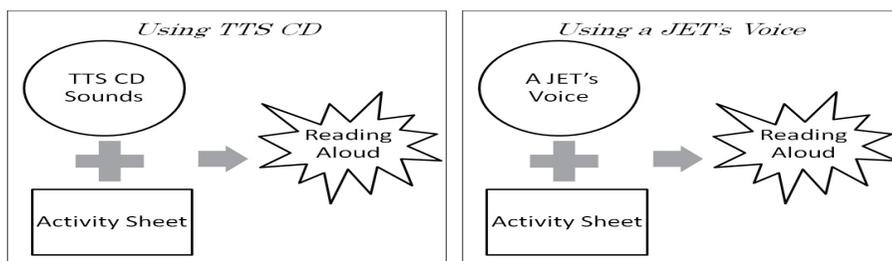


Figure 3. Procedure of Reading Aloud.

The number of repetitions is an important issue for English language learners. Tinkham (1993) found that most learners required five to seven repetitions for learning a group of six paired associates.

To memorize each English sentence in this study, high school students repeated one English sentence twice for one lesson, and repeated one English sentence six times in total.

In order to exclude the influence of gender, both male and female TTS speech sounds were given to the students. This reading aloud task was conducted for five minutes during the six lessons (Table 3).

5.4 Basic Data of the Two Groups

Prior to the experiment, a 10-item cloze test (Appendix A) was administered to Groups A and B to measure their English proficiency. For each correct answer, 1 point in the cloze test was given, with a total score of 10. The preliminary data is summarized in Table 4.

Table 4. Basic Data of Two Groups

	Group A (<i>n</i> = 59)		Group B (<i>n</i> = 56)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Cloze Test	7.08	1.25	7.29	1.66

Note. A correct answer was given 1 point, with a total score of 10.

A two-tailed independent *t*-test⁴ confirmed that a statistically significant difference did not exist between the results of the two groups ($t = -0.74$, $df = 113$, $p = 0.46$, $d = 0.15$). Therefore, they were considered equal in terms of a 10-item cloze test.

The Cronbach's alpha of the cloze test was 0.38. Judging from this score³, the cloze test appears to be difficult for the participants, but we hope they were able to learn the grammatical rule of each English sentence in the cloze test.

5.5 Results and Discussions

5.5.1 Results of a Ten-item Cloze Test

In order to investigate the speech sounds of two kinds of instructions (a) a TTS CD style instruction and (b) a teacher's voice instruction, a 10-item cloze test (Appendix A) was used for the pre-and post-test. The *t*-test was used to compare the results of a ten-item cloze test. The *p* values are generally used to analyze the results of studies. However, the *p* values change according to the sample size. Effect sizes are the absolute value, which are not affected by the sample size. Therefore, the effect sizes are needed to reveal the statistical analyses results (APA, 2009, pp. 32–33; Field, 2009; Mizumoto & Takeuchi, 2008; Mizumoto, 2009). There are two types of effect sizes⁵ (1) *r* and (2) *d* in a *t*-test.

5.5.2 Results of the Reading Aloud Task (1)

In order to measure the development of students' English proficiency using the Reading Aloud Task (1), students in Groups A and B took a post ten-item cloze test. Then, the answers of English sentences No. 1 to 5 in a post ten-item cloze test were selected to analyze the Reading Aloud Task (1). During the Reading Aloud Task (1), students learned English sentences No. 1 to 5 in the cloze test (Appendix A, Appendix B). The students in Group A learned the five English sentences by using an activity sheet (1) (Appendix B) and a TTS CD for three lessons. The students in Group B learned the five English sentences by using an activity sheet (1) (Appendix B) and listening to only a teacher's voice for three lessons.

Table 5 and Figure 4 show the results of items No. 1 to 5 in a post ten-item cloze test.

Table 5. Results of *T*-Test: Items No. 1 to 5

Group	<i>n</i>	Sounds	<i>M</i>	<i>SD</i>
A	59	TTS CD	4.32	0.65
B	56	JET's Voice	4.36	0.74

Note. One point was granted to each correct answer, and the total score was 5.

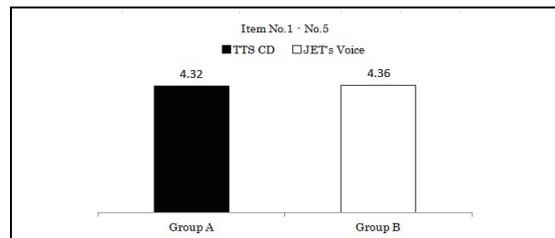


Figure 4. Comparison between TTS CD and the teacher's voice: Items No. 1 to 5.

A two-tailed independent *t*-test was conducted to compare between a TTS CD (Group A) and a teacher's voice (Group B), which were used in the Reading Aloud Task (1).

The results confirmed that a statistically significant difference did not exist ($t = -0.27$, $df = 113$, $p = 0.79$, $d = 0.06$), and the effect size was small. In brief, the result did not show a significant difference between a TTS CD (Group A) and a JET's voice (Group B).

5.5.3 Results of the Reading Aloud Task (2)

In order to measure the development of students' English proficiency using the Reading Aloud Task (2), students in Groups A and B took a post ten-item cloze test. The answers of English sentences No. 6 to 10 in a post ten-item cloze test were selected to analyze the Reading Aloud Task (2). During the Reading Aloud Task (2), students learned English sentences No. 6 to 10 in the cloze test (Appendixes A and B). The students in Group A learned the five English sentences by using an activity sheet (2) (Appendix B) and listening to a teacher's voice, and not a TTS CD, for three lessons. The students in Group B learned the five English sentences by using an activity sheet (2) (Appendix B) and a TTS CD for three lessons. Table 6 and Figure 5 show the results of items No. 6 to 10 in ten-item the post cloze test.

Table 6. Results of *T*-Test: Items No. 6 to 10

Group	<i>n</i>	Sounds	<i>M</i>	<i>SD</i>
A	59	JET's Voice	3.64	1.17
B	56	TTS CD	3.73	1.14

Note. One point was granted to each correct answer, and the total score was 5.

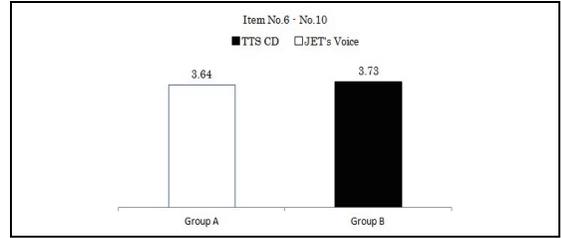


Figure 5. Comparison between TTS CD and the teacher's voice: Items No. 6 to 10.

A two-tailed independent *t*-test was conducted to compare between a teacher's voice (Group A) and a TTS CD (Group B), which were used in the Reading Aloud Task (2).

The results confirmed that a statistically significant difference did not exist ($t = -0.41$, $df = 113$, $p = 0.68$, $d = 0.08$), and the effect size was small. In brief, the result did not show a significant difference between a teacher's voice (Group A) and a TTS CD (Group B).

5.5.4 Results of a Ten-item Cloze Test in Each Group

Table 7 shows the results of a 10-item pre- and post-cloze test in each group. Figure 6 graphically shows the test results of Group A and Group B respectively.

Table 7. Result of Dependent *T*-Test

Group	<i>n</i>	Pre Test		Post Test	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
A	59	7.08	1.25	7.97	1.49
B	56	7.29	1.66	8.09	1.44

Note. One point was given to each correct answer, and the total score was 10.

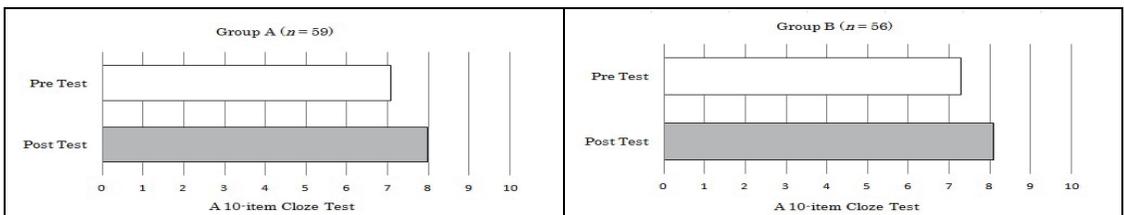


Figure 6. Comparison of pre- and post-test in Groups A and B.

A two-tailed dependent *t*-test confirmed that a statistically significant difference existed between the results of the pre- and post-test in Group A ($t = -5.43$, $df = 58$, $p = 0.00$, $r = 0.58$), and a large effect size was found.

A two-tailed dependent *t*-test also confirmed that a statistically significant difference existed between the results of a pre- and post-test in Group B ($t = -4.20$, $df = 55$, $p = 0.00$, $r = 0.49$), and a middle effect size was found.

The results in Groups A and B showed that TTS speech sounds have enough audio quality as much as a JET's voice, and can be used as audio English language learning materials for

Japanese high school students.

6. SUMMARY AND CONCLUDING REMARKS

This study focuses on the use of TTS synthesized speech sounds introduced in English grammar instruction in high school education.

There are two limitations in our experiment.

1. The scope of our present study was limited to the question sentence level.
2. This study only compared the usability of TTS speech sounds and a Japanese-English language teacher's voice.

Given the research design of this study, some of our conclusions remain tentative. To overcome this problem, future studies on (1) a variety of sentences and (2) a comparison between TTS speech sounds and the ALT's voice, are needed.

Research question (RQ) in this study was "Is there any significant difference in basic English verb learning between the two kinds of reading aloud instructions: (a) a TTS CD style instruction, and (b) a verbal instruction made by a Japanese English teacher?"

Our findings revealed two set of results.

First, the results obtained in a pre- and post-cloze test show that a statistically significant difference existed in Group A and Group B respectively (see section 5.5.4). This suggests that TTS speech sounds can be used as audio English language learning materials for Japanese high school EFL education.

Second, a statistically significant difference did not exist between the two types of instructions: (a) a TTS CD style instruction, and (b) a JET's voice instruction (see sections 5.5.2 and 5.5.3).

The results confirm that TTS speech sounds used in this study had enough usability as speech sounds for Japanese high school EFL education. The difference or gap did not appear between the TTS speech sounds and the teacher's voice in this study.

Our results prove the usability of TTS speech sounds in high school EFL education. Without the aid of ALTs by using the TTS system, high school English teachers can easily create audio materials to fit their students' English proficiency.

English speech sounds can be of great help for Japanese EFL learners to study English (Hatori, 1977; Kohno, 1984; Palmer, 1921). However, almost all of the audio language learning materials used in Japanese high schools are developed by commercial publication companies.

We Japanese-English teachers sometimes feel that such audio materials do not fit our students' English proficiency. Using TTS will be one way for teachers who want to improve audio materials for their students during their lessons. We believe that teachers can use TTS synthesized digital speech sounds so that they can create audio English language learning materials suitable for their students, based on their English proficiency level. For teachers to create audio English learning materials for high school students, students will need to find

opportunities to listen to English speech sounds in order to develop their English proficiency. Their teachers can help develop their students' English proficiency little by little, by using TTS speech sounds during their lessons.

We hope that a greater number of English teachers in Japanese high schools use TTS speech sounds to create audio materials for their lessons and that additional research on TTS synthetic speech is carried out to further develop high school EFL education.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the cooperation and support of the high school teachers and students to take time for this research. We would like to express our deep gratitude to Dr. Shigeru Yamane, Kansai University for his valuable comments on this paper. We also thank Dr. Eiichi Yamamoto, Professor Taichi Usami, and Professor Takaichi Okuda, Kansai University for their warm encouragement to our high school English educational studies.

NOTES

1. We received permission through the selling agent of an audio database in this study to create TTS CD-ROMs.
2. Mr. Kenji Otsuka has been teaching English in public junior high schools in Hokkaido since 1986, and he was awarded as a professional English teacher by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in 2010.
3. According to Pallant (2007, p. 6, p. 95), Cronbach's alpha of a scale should be above 0.70.
4. A two-tailed independent *t*-test: Effect size $d = 0.20$, small effect size (S); $d = 0.50$, medium effect size (M); $d = 0.80$, large effect size (L).
5. A two-tailed dependent *t*-test: Effect size $r = 0.10$, small effect size (S); $r = 0.30$, medium effect size (M); $r = 0.50$, large effect size (L).

REFERENCES

- American psychological association. (2009). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.
- Azuma, J. (2008). Applying TTS technology to foreign language teaching. In F. Zhang & B. Barber (Eds.), *Handbook of research on computer-enhanced language acquisition and learning* (pp. 497-506). New York: Information Science Reference.
- Chew, L. C. (2009) Promises and challenges of e-assessments: A case of multimedia listening testing. *LET Kansai Chapter Collected Papers, 12*, 1–19.
- Dutoit, T. (1997). *An introduction to text-to-speech synthesis*. Dordrecht; Netherland: Kluwer Academic Publishers.
- Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Los Angeles: SAGE.
- Hatori, H. (1977). *Eigokyouiku no shinrigaku [Psychology in English language education]*.

Tokyo: Taishuukanshoten.

- Hirai, A., & O'ki, T. (2011). Comprehensibility and naturalness of Text-To-Speech synthetic materials for EFL listeners. *JACET Journal*, 53, 1–17.
- Ikeda, M., & Takeuchi, O. (2003). Ondoku ni yoru sougouteki na eigonouryoku no sokutei: Jissoutekikenkyu [Reading aloud: An indicator of integrative ability of EFL]. *LET Kansai Chapter Collected Papers*, 9, 23–30.
- Ishihara, Y. (2002). Koukou Reading no jugyou niokeru ondoku no kouka ni tsuitenno ichikousatsu. [An empirical study on the effectiveness of reading aloud in English language reading classes at a senior high school level]. *Hiroshima Daigaku Huzoku Cyuu-koutougakkou kiyuu*, 49, 71-75.
- Jones, C., Berry, L. & Stevens, C. (2007). Synthesized speech intelligibility and persuasion: Speech rate and non-native listeners. *Computer Speech and Language*, 21, 641–651.
- Kataoka, H. (2009). Text-To-Speech (TTS) synthesis technology wo katsuyoushita eigo kyouikukyousai no kaihatsu to nihonjin no onseininshiki [The use of Text-To-Speech (TTS) synthesis technology for English education: Speech recognition of Japanese EFL learners]. *Journal of Kansai University Graduate School of Foreign Language Education and Research*, 7, 1–33.
- Kitamura, E. (2007). Ondoku hissha jikan to koukousei no eigo nouryoku no kankei [Relationship between writing while reading aloud and English proficiency of high school students]. *STEP Bulletin*, 17, 81–94.
- Kohn, M. (1984). *Eigojugyou no kaizou [Remodeling of English language lessons]*. Tokyo: Tokyoshoseki.
- Ministry of Education, Culture, Sports, Science and Technology. (2010). Kyouka shidouniokeru ICT katsuyou. ([ICT practical use in course instruction].) from http://www.mext.go.jp/a_menu/shotou/zyouhou/1259413.htm (Accessed 2010.10. 30).
- Miyasako, N. (2002). Koukousei no ondoku to eigoryoku wa kankei ga aruka [Does oral reading ability have a relationship with English language ability for senior high school students?]. *STEP Bulletin*, 14, 14–25.
- Miyasako, N. (2007). A theoretical and empirical approach to oral reading. *Language Education & Technology*, 44, 135–154.
- Mizumoto, A. & Takeuchi, O. (2008). Kenkyuuronbun ni okeru koukaryou no houkoku no tameni [Basics and considerations for reporting effect sizes in research papers]. *Studies in English Language Teaching*, 31, 57–66.
- Mizumoto, A. (2009). Stats Stuff. <http://www.mizumot.com/stats.html> (Accessed 2009.05.09)
- Otsuka, K. (2009). Easy Questions. <http://ok.ko.kg/> (Accessed 2009.08.10)
- Pallant, J. (2007). *SPSS survival manual: a step by step guide to data analysis using SPSS for Windows* (3rd ed.). Maidenhead: Open University Press.
- Palmer, H. E. (1921) The principles of language study. In the Institute for Research in Language Teaching (Ed.), *The selected writings of Harold E. Palmer: Pāmā senshū*,

- Dai 1 Kan, (1999), (pp. 331–520). Tokyo: Hon-no Tomosha.
- Suzuki, J. (1998). Ondokusidou-saihyouka: Ondokushidou no kouka ni kansuru jissyoutekikikenkyu [A reappraisal of reading aloud: An empirical study on the effectiveness of reading aloud]. *LET Kansai Chapter Collected Papers*, 7, 14–28.
- Takeuchi, O. (2003). *Yoriyoi gaikokugo gakushuuhou o motomete [Synopsis in English of Searching for better language learning strategies: Studies on good language learners in the Japanese FL context]*. Tokyo: Shohakusha.
- Tinkham, T. (1993). The effects of semantic clustering on the learning of second language vocabulary. *System*, 21, 371–380.
- Watanabe, H. (1988). Choukai to dokkai no setten [The contact point between listening and reading]. *Leo*, 17, 41–77.
- Watanabe, H. (2009). Sokudoku ni tsunagaru ondoku no susume: Ondoku mokudoku no hayasa wo dou toraeruka [Recommendation for reading aloud which leads to reading fast: how to recognize the speed of silent reading and/or reading aloud]. *The English Teachers' Magazine*, 58, 9, 23–25.
- Widdowson, H. G. (1978). *Teaching English as communication*. Oxford; New York: Oxford University Press.
- Yoshida, S. (2008). Development and practice of an electronic phrasal verb wordbook with GIF animations. *Paper presented at the Proceedings of the 3rd International Conference of World CALL 2008*, 84.

Appendix A. Ten English Sentences in the Cloze Test

1 次の英文の () 内に適切な語を記号とつづりで答えなさい。			
解答例 ア) was			
1. () you a student?	ア) Was	イ) Is	ウ) Am
2. () you from Sapporo?	ア) Is	イ) Was	ウ) Are
3. () this a book?	ア) Am	イ) Is	ウ) Are
4. () Tokyo Disney Land in Tokyo?	ア) Is	イ) Am	ウ) Were
5. () there a pen on the desk?	ア) Is	イ) Was	ウ) Are
6. Is Japanese () in China?	ア) speak	イ) spoke	ウ) spoken
7. () there any good restaurants in your city?	ア) Are	イ) Is	ウ) Am
8. () you have any good restaurants in your city?	ア) Are	イ) Do	ウ) Does
9. () you like English?	ア) Do	イ) Does	ウ) Did
10. () you study English every day?	ア) Did	イ) Do	ウ) Are

Appendix B. Activity Sheets (1) and (2)

月 日 () 実施 幸福文法 初級編 Activity Sheet (1)		月 日 () 幸福文法 初級編 Activity Sheet (2)	
声に出しながら、暗記しましょう。		声に出しながら、暗記しましょう。	
1. Are you a student?	あなたは学生ですか。	6. Is Japanese spoken in China?	日本語は中国で話されていますか。
Yes, I am.	はい、そうです。	No, it isn't.	いいえ、話されていません。
2. Are you from Osaka?	あなたは大阪出身ですか。	7. Are there any good restaurants in your city?	あなたの町にはよいレストランがありますか。
Yes, I am.	はい、そうです。	Yes, there are.	はい、あります。
3. Is this a book?	これは本ですか。	8. Do you have any good restaurants in your city?	あなたの町にはよいレストランがありますか。
No, it isn't.	いいえ、本がいません。	No, we don't.	いいえ、ありません。
4. Is Tokyo Disneyland in Tokyo?	東京ディズニーランドは東京にありますか。	9. Do you like English?	あなたは英語が好きですか。
No, it isn't. It's in Chiba Prefecture.	いいえ、千葉にあります。	Yes, I do.	はい、好きです。
5. Is there a pen on the desk?	机の上にペンはありますか。	10. Do you study English every day?	あなたは毎日英語を勉強しますか。
Yes, there is.	はい、あります。	No, I don't.	いいえ、しません。