MAXIMUM LISTENING SPEEDS FOR THE BLIND

Chieko Asakawa, Hironobu Takagi*

IBM Japan, Tokyo Research Lab.
1623-14 Shimotsuruma, Yamato-shi
Kanagawa, 242-8502 JAPAN
{chie, takagih}@jp.ibm.com

Shuichi Ino, Tohru Ifukube**

*University of Tokyo
4-6-1 Komaba, Meguro-ku,
Tokyo 153-8904 JAPAN
Ifukube@rcast.u-tokyo.ac.jp

ABSTRACT

Blind people usually use voice output using a computer, however, there is little objective data about how fast or accurately they can obtain information in a fixed amount of time In this paper, we describe the highest and the most suitable listening rate for the blind based on our human factors experiments, aiming at producing a kind of indicator for use by developers.

We experimented with the highest and the most suitable listening rates for blind users with objective and subjective test methods. The results showed that the advanced blind testers could listen to the spoken material at speeds 1.6 times faster than the highest rate of the tested TTS (Text-to-Speech) engine. This indicates that the currently available TTS engines should support faster rates. It also showed that the highest rate often changes depending on the difficulty of the sentences and words. These results would be valuable and useful indicators for developers to design applications for the blind and to improve the nonvisual user interfaces.

1. INTRODUCTION

A computer is an essential tool for the blind as a substitute for their eyes, and it enables them to access printed documents. Applications to support blind users [1, 2, 3] have been rapidly increasing and have spread all over the world. Such applications mostly use voice output, but there is little objective data about how quickly and accurately users can understand a document by listening to it, or about the most suitable listening rates._The initial rate and the highest rate of such applications often depend on the developer's subjective decisions. Therefore advanced users (users who are experienced with computers) feel significant dissatisfaction with the reading rates. The reading rate is one of the most important issues to improve the voice usability for persons with visually impairments who need to access the information using voice output, one dimensionally.

We performed experiments with blind subjects to measure their most comfortable listening rate and how fast and accurately they can obtain information in a fixed amount of time [4].

Related work has reported how well senior citizens and hearing-impaired people can recognize voice information by slowing down the reading rates [5, 6]. This research was characterized by investigating blind users' listening or recall abilities while speeding up the reading rates as much as possible.

In this research, we experimented with the highest and the most suitable listening rates using subjective and objective methods. Our results showed that the average highest rates from subjective evaluations were approximately equal to the averages from the objective evaluations. The advanced blind testers were able to listen to the spoken material at rates much faster than we had anticipated. It was also shown that the listening rate for each subject is very much influenced by the difficulty of the sentences and words. These results will be very useful indicators for developers who design and develop assistive technology software. In this paper, we will describe the experiments and discuss the experimental results, and then future plans will follow.

2. THE HIGHEST AND THE MOST SUITABLE LISTENING RATES

2.1. Objectives

The experimental objective is to measure the highest listening rate that blind users can recognize and the most suitable listening rate for blind users to listen to. The highest rate here means that the rate such that they can recognize about 50 percent of the content of the information. The most suitable listening rate here means the rate such that they can recognize about 100 percent of the content of the information comfortably.

We investigated the following questions:

1. Are the highest and the most suitable listening rates different in subjective and objective evaluations?
2. Does the computer experience of the subjects affect their highest and most suitable listening rates?
3. Will a repeated presentation method help to improve their highest and most suitable listening rates?
4. Will it reduce the recognition rate when there are unfamiliar words such as terms used only by experts?
5. Is the recognition ability affected by their knowledge and experience apart from their computer experience?

2.1.1. Methods

This experiment consisted of a subjective evaluation and an objective evaluation, both using short sentences. For each evaluation method, a repeated presentation test (or method) and a random presentation test (or method) was done. One set of test data consisted of 15 wave files that are recorded at different rates.

In the subjective evaluation, the subjects were asked to report their own highest and most suitable rates for each data set. After the experiment was over, the average highest and the average most suitable rate for each subject were defined.

In the objective evaluation, the subjects were asked to recall what they heard. After the experiment was over, we transcribed the recorded voices and the recall rate (RR) was defined as the number of correct words compared with the total number of words in the test data. When the RR first became higher than 90%, this was recognized as the most suitable rate. The accuracy level of 50% was regarded as the highest rate. (See figure 1)

In the repeated presentation test, one set of test data consisted of the same sentence at different rates, while the random presentation consisted of different sentences at different rates. Both presentation methods were used to evaluate how the highest and the most suitable rates would be affected by listening to the same sentences repeatedly.

2.1.2. Subjects

Seven legally blind subjects who are experienced with computers participated in the experiments (two advanced, two intermediate, and three novice users). Three were in their 20s, and one came from each age bracket of the 30s, 40s, and 60s. Three were female, and four were male.

2.1.3. Test data

A recorded human voice was used in order not to be affected by the quality differences among TTS systems. The original data was recoded at 180 wpm. One set of test data consisted of 15 wave files, recoded and then adjusted to the different rates shown in Table 1 using the time stretch and compression functions of Cool Edit [7]. The file format was windows PCM (.wav, 22 kHz and 16-bit sampling monaural).

The test data (sentences) were selected from the ATR Speech Database [8]. As much as possible, we tried to select relatively short sentences with familiar words, such as “It is delicious to have a cup of Russian tea with plenty of jam,” “I was looking for someone who would protect me,” and so on. Since the subjects were Japanese, Japanese sentences were used for the entire experiment (and some may seem a bit odd when translated into English). We converted the measurements to morae per minute, so that the measured Japanese rates correspond to the English rates in words per minute.

Speed		Morae/min. (in Japanese)	TTS (Japanese)	Words/min. (in English)
1	x3.6	1,755		1,755
2	x3.4	1,658		1,658
3	x3.2	1,560		1,560
4	x3.0	1,463		1,463
5	x2.8	1,365		1,365
6	x2.6	1,268		1,268
7	x2.4	1,170		1,170
8	x2.2	1,073		1,073
9	x2.0	975		975
10	x1.8	878	highest (862)	878
11	x1.6	780		780
12	x1.4	683		683
13	x1.2	585		585
14	x1.0	488	default (517)	488
15	x0.8	390		390

Table 1. *List of Speech Rates.*

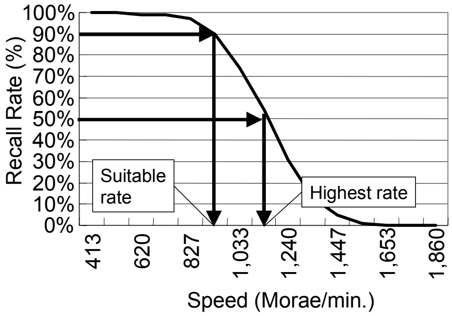


Figure 1. *Definition of the Highest and Most Suitable Rates.*

2.1.4. Apparatus

Hardware:

Two laptop computers: one for presenting the test data and one for writing notes. There was a speaker for the experimenters and a pair of headphones for the subject. An MD disk was used for recoding the subjects’ recall in the objective evaluation.

Software:

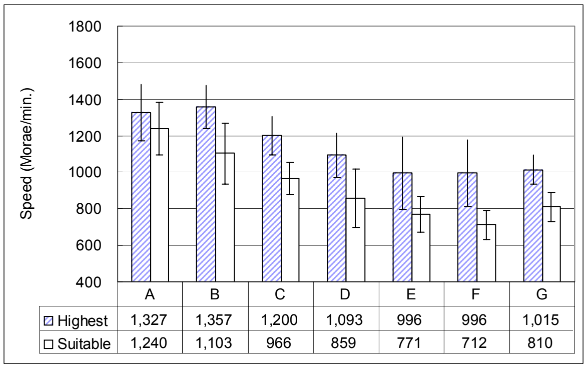
Cool Edit was used for the time compression of the recorded sound data. The sequence of the sound data was described using XML and played by using Internet Explorer with Real Player_

2.2. Subjective evaluation experiment

2.2.1. The procedure of the experiment

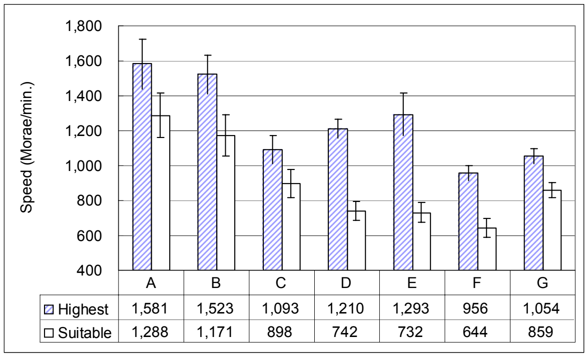
Subjects were given an overview of the experiment for ten minutes before the experiment began. The definition of the highest and the most suitable listening rates were explained in detail. They were asked to report the highest rate when they recognized about 50 percent of the sentence, but not all of it, and they were asked to report the most suitable rate when they recognized the entire sentence without listening effort. The wave file at each rate is only presented once. After selecting a rate, the next slower version was still presented, and the test subject was allowed to change to the slower recording as the highest or the most suitable rate, if desired. After receiving the instructions, three practice data sets for training were used in the same way as the real experiment. During the training, the subjects were allowed to ask questions.

The repeated presentation test was performed first. Ten sets of data were used. The fastest wave file in a set was presented first, and then the next slower one was presented, and so on, in the same order as shown in Table 1. When the subject first subjectively recognized the sentence, that rate was reported as the highest rate for that subject. After that, the rate continued to be reduced, and when the subject felt that it was the most suitable rate, that rate was recorded. Each wave file was presented once. Subjects listened to the data with the headphones and the experimenters monitored the test with the speaker. After the repeated presentation test was over, the random presentation test for five data sets was performed by controlling the decreasing rates in the same way as in the repeated presentation, but while using a different (randomly selected) sentence at each rate. The total time for each subjective evaluation was 40 minutes, including the instructions and training.



	A	B	C	D	E	F	G
▨ Highest	1,327	1,357	1,200	1,093	996	996	1,015
□ Suitable	1,240	1,103	966	859	771	712	810

Figure 2. Results of Subjective Evaluation Test (Repeated Presentation).



	A	B	C	D	E	F	G
▨ Highest	1,581	1,523	1,093	1,210	1,293	956	1,054
□ Suitable	1,288	1,171	898	742	732	644	859

Figure 3. Results of Subjective Evaluation Test (Random Presentation).

2.3. Objective evaluation experiment

2.3.1. The procedure of the experiment

In the objective evaluation, the short sentences are presented to the subjects and they are asked to recall what they have heard. The highest and the most suitable listening rates are defined based on their recall rate (RR). The test data was presented in the same way as in the previous experiment, that is, the fastest wave file in a set was presented first, and then the next slower one was presented, and so on. One data set was completed when the subject recalled all of the sentence (RR is 100 percent) or when all 15 wave files had been presented, and then the next set was started. For the test data, we provided ten sets for the repeated presentation method, and 5 sets for the random presentation. In total, it took 45 minutes for the objective evaluation, including the brief instructions and three data sets for training.

2.3.2. Determination of the highest and the most suitable listening rates

After the experiment was over, we transcribed the recorded voices and the recall rate (RR) was defined as the number of correct words compared with the total number of words in the test data. In the case of “This is a test” (4 words), if the number of correct words was 3, the RR would be 75%. After the average recall rate for each rate for each test subject was determined, we defined the highest and the most suitable rates. When the RR first reached more than 90%, this was recognized as the most suitable rate. The accuracy level of 50% was regarded as the highest rate.

3. RESULTS

3.1. Subjective evaluation

Figure 2 shows the results for the repeated presentation test and Figure 3 is for the random presentation test. Subjects A and B are advanced users, C and D are intermediate, and E, F, and G are novice users. This shows that the highest and the most suitable rates for the advanced users are always faster than those of the other users. But even the novice users’ average suitable rate is 1.4 times faster or 1.6 times faster than the default rate shown in Table 1. Comparing the repeated and random presentation tests, the highest rate was much faster with the random presentation, which was not in accord with our prediction. However, the most suitable rate was almost the same between two tests.

3.2. The objective evaluation

Figure 4 shows the average RR for each rate in the repeated presentation test. The horizontal bar indicates the rate and the vertical bar shows the RR. The results fit a sigmoid curve as shown in the figure. The advanced subjects always show higher RRs at higher reading rates compared to the others. Observing the two advanced subjects, B always has a significantly higher RR, while A sometimes has a lower RR at the lower reading rates (less than 3.0), almost the same as the intermediate subjects and off of the sigmoid curve. The highest rate was almost equal to each other, the subjective and objective evaluations. However, the most suitable rate was lower in the objective evaluation.

Figure 5 shows the results for the random presentation test. These RR results do not fit a curve. Therefore, it was hard to define the highest and the most suitable listening rates with this test method.

4. DISCUSSION

The highest rates for advanced users were 2.6 times faster and 2.8 times faster than the default rate of the tested TTS engine. This indicates that they could understand at least 50 percent of the information at the rate of 1,300 Mora/min. or 500 wpm. This is quite fast and beyond our expectations. This indicates that advanced blind users have outstanding listening ability, which as hard to predict, since in Japan the frequently used TTS systems only support rates under 900 Morae.

The novice subjects reported that they usually used the default rate of the tested TTS engine when using a screen reader. However, the average rate of their most suitable listening rate was about 1.6 times faster than the default rate. It indicates that it might be possible to improve their working environments by changing the default rate, even without any new settings.

Comparing the repeated presentation and the random presentation in the subjective evaluation, we predicted the highest and the most suitable rates would be higher with the repeated presentation, however, there was no significant

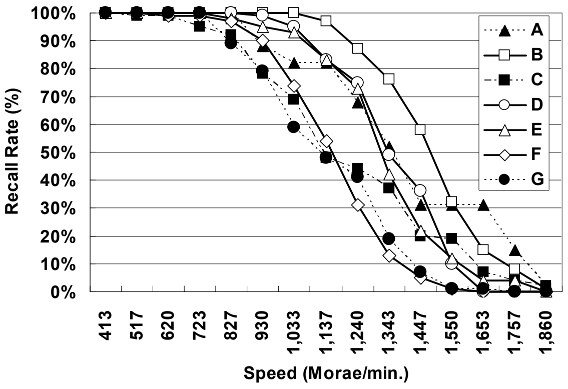


Figure 4. Results of Objective Evaluation Test (Repeated Presentation).

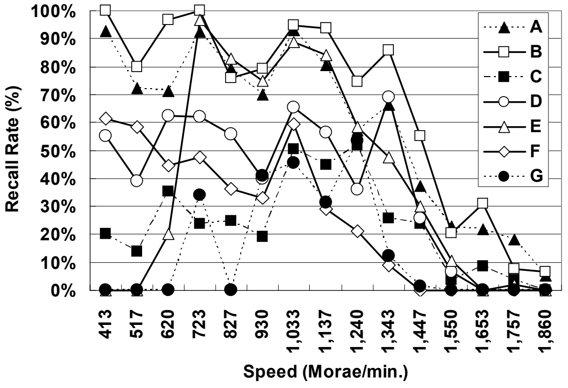


Figure 5. Results of Objective Evaluation Test (Random Presentation)

difference between them. However we could not define both rates using the random presentation for objective evaluation. It can be concluded that the random presentation test would not be needed in follow-up experiments.

Finally, from this experiment it was clear that the RR is different depending on the difficulty of the sentences and the difficulty is different for each experimental subject. From their comments, it is clear that there are certain conditions for each listener when they feel it is difficult to listen to something. For example some commented that it is confusing when there are similar words in one sentence or when there are unfamiliar words in one sentence, while this condition posed little difficulty to others. They often encounter these difficulties when they access information by listening, and they want to slow down the rate in such conditions. However, they usually do not do so, since currently, they can only change the rate using the menu or keyboard commands. It might be suggesting that the information read at the higher rate might not be understood correctly and some words might be lost. However slowing the rate will cause stress to advanced users.

5. CONCLUSIONS

In this paper, we discussed reading rates, a key interface factor when using a computer to improve the voice usability for the blind. With the currently available TTS systems, the default rate or the highest rate is dependent on the developers' subjective decision. We therefore investigated the highest and the most suitable listening rates for blind subjects. The result shows that the highest rates of the advanced subjects reached 1,300 Mora per minute, which is quite rapid. In Japan the frequently used TTS systems only support rates under 900 Mora. It indicates that such systems need high quality support for faster rates for blind users. The results also indicate that supporting faster rates would not by itself solve all of the issues in voice usability. The RR changes based on the difficulty of the sentences, and these changes are often caused by smaller units of the sentence, such as words and phrases. In addition, the perceived difficulty varied for each of our test subjects. This indicates that the reading rates should be easily and interactively changed by the users with immediate response.

In the future, we would like to extend these results as a kind of indicator for developers of assistive technology software. New interfaces should be developed to solve such issues, allowing blind people to control the reading rates flexibly in order to improve voice usability.

6. REFERENCE

JAWS, Freedom Scientific Inc., http://www.freedomscientific.com/

Asakawa, Itoh, "User Interface of a Non-visual Web Access System", IPSJ (Information Processing Society of Japan) Journal, Vol.40, No.2, pp.453-459, 1999 (in Japanese)

T. Watanabe, "Study for GUI-based Screen Reader Using the Voice Output", Doctoral Dissertation of Hokkaido University, 2001

C. Asakawa, H. Takagi, S. Ino, T. Ifukube, "The Highest and the Most Suitable Listening Rate for the Blind in the Screen Reading Process", in Proceedings of Human Interface Symposium 2002, Human Interface Society in Japan, 2002 (in Japanese)

T. Ifukube, "Speech Signal Transformations for the Elderly Hearing Impaired", Vol. 84, No. 5, pp.325-328, 2001_(in Japanese)

N. Seiyama, et al., "Development of A High-Quality Real-Time Speech Rate Conversion System", Transaction on IEICE, D-II.Vol.J84-D-II. No.6. pp.918-926, 2001_(in Japanese)

CoolEdit, Syntrillium, http://syntrillium.com/

ATR Speech Database (in Japanese), http://www.red.atr.co.jp/detabase.html