

A SENTENCE ANALYSIS METHOD FOR A JAPANESE BOOK READING MACHINE FOR THE BLIND

Yutaka Ohyama, Toshikazu Fukushima, Tomoki Shutoh and Masamichi Shutoh
C&C Systems Research Laboratories
NEC Corporation
1-1, Miyazaki 4-chome, Miyamae-ku,
Kawasaki-city, Kanagawa 213, Japan

ABSTRACT

The following proposal is for a Japanese sentence analysis method to be used in a Japanese book reading machine. This method is designed to allow for several candidates in case of ambiguous characters. Each sentence is analyzed to compose a data structure by defining the relationship between words and phrases. This structure (named *network structure*) involves all possible combinations of syntactically collect phrases. After *network structure* has been completed, heuristic rules are applied in order to determine the most probable way to arrange the phrases and thus organize the best sentence. All information about each sentence —— the pronunciation of each word with its accent and the structure of phrases —— will be used during speech synthesis. Experiment results reveal: 99.1% of all characters were given their correct pronunciation. Using several recognized character candidates is more efficient than only using first ranked characters as the input for sentence analysis. Also this facility increases the efficiency of the book reading machine in that it enables the user to select other ways to organize sentences.

1. Introduction

English text-to-speech conversion technology has substantially progressed through massive research (e.g., Allen 1973, 1976, 1986; Klatt 1982, 1986). A book reading machine for the blind is a typical use for text-to-speech technology in the welfare field (Allen 1973). According to the Kurzweil Reading Machine Update (1985), the Machine is in use by thousands of people in over 500 locations worldwide.

In the case of Japanese, however, due to the complexities of the language, Japanese text-to-speech conversion technology hasn't progressed as fast as that of English. Recently a Japanese text-to-speech synthesizer has been introduced (Kabeya et al. 1985). However, this synthesizer accepts only Japanese character code strings and doesn't include the character recognition facility.

Since 1982, the authors have been engaged in the research and development of a Japanese sentence analysis method to be used in a book reading machine for the blind. The first version of the Japanese book reading machine, which is aimed to examine algorithms and its performance, has developed in 1984 (Tsuji and Asai 1985; Tsukumo and Asai 1985; Fukushima et al. 1985; Mitome and Fushikida 1985, 1986). Figure 1 shows the book reading process of the machine. A pocket-size book is first scanned, then each character on the page is detected and recognized. Sentence analysis (parsing) is accomplished by using character recognition result. Finally, synthesized speech is generated. The speech can be recorded for future use. The pages will turn automatically.

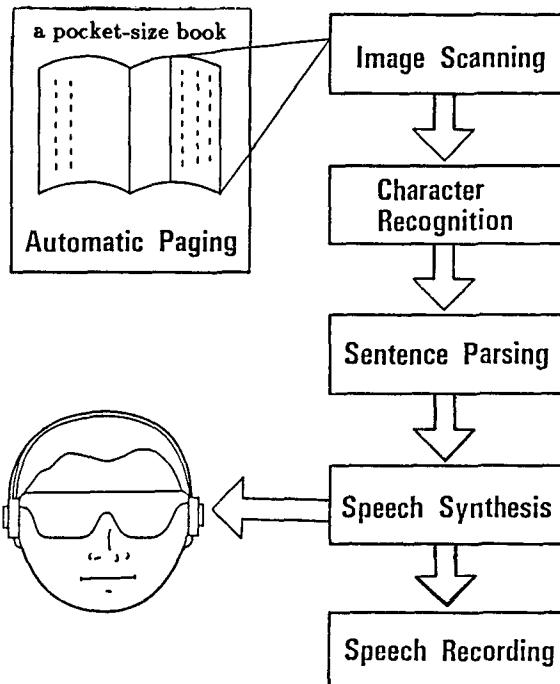


Figure 1. The Book Reading Machine Outline.

The Japanese sentence analysis method that the authors have developed has two functions: One, to choose an appropriate character among several input character candidates when the character recognition result is ambiguous. Two, to convert the written character strings into phonetic symbols. The written character strings are made up Kanji (Chinese) characters and kana (Japanese consonant-vowel combination) characters. These phonetic symbols depict both the pronunciation and accent of each word. The structure of the phrases is also obtained in order to determine the pause positions and intonation.

After briefly describing the difficulty of Japanese sentence analysis technology compared to that of English, this paper will outline the Japanese sentence analysis method, as well as experimental results.

2. Comparison of Japanese and English as Input for a Book Reading Machine

In this section, the difficulty of Japanese sentence analysis is described by comparing with that of English.

2.1 Conversion from Written Characters to Phonetic Symbols

In English, text-to-speech conversion can be achieved by applying general rules. For exceptional words which are outside the rules, an exceptional word dictionary is used. Accentuation can be also achieved by rules and an exceptional dictionary.

Roughly speaking, Japanese text-to-speech conversion is similar to that of English. However, in case of Japanese, more diligent analysis is required. Japanese sentences are written by using Kanji characters and kana characters. Thousands of kinds of Kanji characters are generally used in Japanese sentences. And, most of the Kanji characters have several readings (Figure 2 (a)). On the other hand, the number of kana characters is less than one hundred. Each kana character corresponds to certain monosyllable. Therefore, in the conversion of kana characters, kana-to-phoneme conversion rules seem to be successfully applied. However, in two cases, kana characters は and ん are used as Kaku-Joshi, Japanese preposition which follows a noun to form a noun phrase, then the pronunciation changes (Figure 2 (b)). Subsequently the reading of numerical words also changes (Figure 2 (c)).

As described above, the pronunciation of each character in Japanese sentences is determined by a neighbor character which combines to form a word. There are too many exceptions in Japanese to create general rules. Therefore, a large size word dictionary which covers all commonly used words is generally used to analyze Japanese sentences.

2.2 Required Sentence Analysis Level

In English sentences, the boundaries between words are indicated by spaces and punctuation marks. This is quite helpful in detecting phrase structure, which is used to determinate pause positions and intonation.

On the contrary, Japanese sentences only have punctuation marks. They don't have any spaces which indicate word boundaries. Therefore, more precise analysis is required in order to detect word boundaries at first. The structure of the sentence will be analyzed after the word detection.

日	<u>hi</u>	(day / sun)
日本	<u>ni-hon</u> <u>nip-pon</u>	(Japan) (Japan)
日時	<u>nichi-ji</u>	(date and time)
日下	<u>kusa-ka</u>	(a Japanese last name)
月日	<u>gap-pi</u> <u>tsuki-hi</u>	(date) (months and days)
今日	<u>kyo-u</u> <u>kon-nichi</u>	(today) (recent days)
一日	<u>ichi-nichi</u> <u>ichi-jitsu</u> <u>tsui-tachi</u>	(one day) (one day) (the 1st day of a month)
二日	<u>futsu-ka</u>	(the 2nd day of a month / two days)

(a) Kanji Characters

はなは	きれいだ	(Flowers are beautiful.)
<u>ha-na-wa</u>	<u>ki-re-i-da</u>	
へやへ	はいる	(Entering the room.)
<u>he-ya-e</u>	<u>ha-i-ru</u>	

(b) Kana Characters

一本	<u>ip-pon</u>	(one [pen, stick,...])
二本	<u>ni-hon</u>	(two [pens, sticks,...])
三本	<u>san-bon</u>	(three [pens, sticks,...])

(c) Numerical Words

Figure 2. Examples of Japanese Word.

2.3 Character Recognition Accuracy

English sentences consist of twenty-six alphabet characters and other characters, such as numbers and punctuations. Because of the fewer number of the English alphabet characters, characters can be recognized accurately.

Japanese sentences consist of thousands of Kanji characters, more than one hundred different kana characters (two kana character sets —— Hiragana and Katakana —— are used in Japanese sentences) and alphanumeric characters. Because of the variety of characters, even when using a well-established character recognition method, the result is sometimes ambiguous.

3. Characteristics of Sentence Analysis Method

The Japanese sentence analysis method has the following characteristics.

1. The mixed Kanji-kana strings are analyzed both through word extraction and syntactical examination. An internal data structure (named *network structure* in this paper), which defines the relationship of all possible words and phrases, is composed through word extraction and syntactical examination. After *network structure* has been completed, heuristic rules are applied in order to determine the most probable way to arrange the phrases and thus organize a sentence.
2. When an obtained character recognition result is ambiguous, several candidates per character are accepted. Unsuitable character candidates are eliminated through sentence analysis.

3. Each punctuation mark is used as a delimiter.

Sentence analysis of Japanese reads back to front between punctuation marks. For example, the analysis starts from the position of the first punctuation mark and works to the beginning of the sentence. Thus, word dictionaries and their indexes have been organized so they can be used through this sequence.

4. The sentence analysis method is required for short computing time to analyze unrestricted Japanese text. Therefore, it has been designed not to analyze deep sentence structure, such as semantic or pragmatic correlates.

5. By the user's request, the book reading machine can read the same sentence again and again. If the user wants to change the way of reading (e.g. in the case that there are homographs), the machine can also create other ways of reading. In order to achieve this goal, several pages of sentence analysis result is kept while the machine is in use.

4. Outline of Sentence Analysis System

As shown in Figure 3, the Japanese sentence analysis system consists of two subsystems and word dictionaries. Two subsystems are named "network structure composition subsystem" and "speech information organization subsystem", respectively. These subsystems work asynchronously.

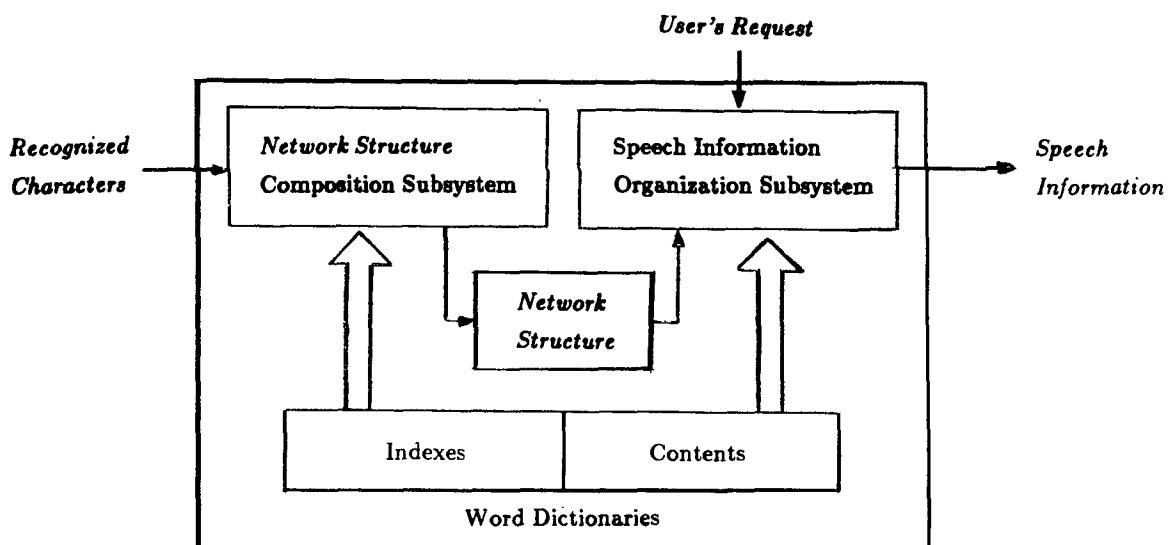


Figure 3. Sentence Analysis System Outline.

4.1 Network Structure Composition Subsystem

As the input, the *network structure* composition subsystem receives character recognition results. When the character recognition result is ambiguous, several character candidates appear. During the character recognition, the probability of each character candidate is also obtained. Figure 4 is an example of character recognition result. Figure 4 describes: The first character of the sentence as having three character candidates. The fifth and seventh characters as having two candidates. Except the fifth character, all of the first ranking character candidates are correct. However, the fifth character proves an exception with the second ranking character candidate as the desired character.

With the recognized result, the *network structure* composition subsystem is activated. Figure 5 describes how the recognition result (shown in Figure 4) is analyzed.

Through the detection of punctuation marks in the input sentence (recognition result), the subsystem determines the region to be analyzed. After one region has been analyzed, the next punctuation mark which determines the next region is detected. In case of Figure 5, for example, whole data will be analyzed at once, because the first punctuation mark is located at the end of the sentence.

Characters in the region are analyzed from the detected punctuation to the beginning of the sentence. The analysis is accomplished by both word extraction and syntactical examination. Words in dictionaries are extracted by using character strings which are obtained by combining character candidates. The type of the characters (kana, Kanji etc.) determines which index for the dictionaries will be used.

Input Text

文章を解析する。

(Analyze a sentence.)



	1	2	3	4	5	6	7	8
1st Candidate	文	章	を	解	折	す	る	。
2nd Candidate	丈				折		る	
3rd Candidate	又							

Figure 4. Character Recognition Result Example.

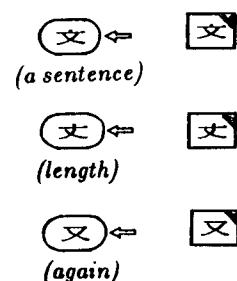
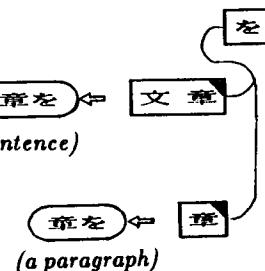
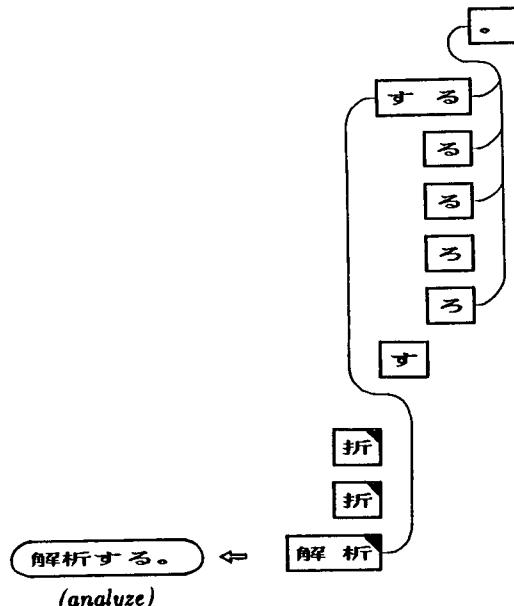
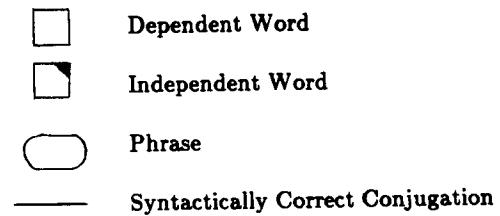


Figure 5. Sentence Analysis Example.

After extracting the words, phrases are composed by combining the words. Using syntactical rules (i.e. conjugation rules), only syntactically correct phrases are composed.

Finally, by using these phrases, *network structure* is composed. *Network structure* obtained through the analysis described in Figure 5 is shown in Figure 6. This structure involves the following information.

- hierarchical relationship between sentence, phrases and words
- syntactical meaning of each word
- pointers to the pronunciation and accent information of for each word in dictionaries
- pointers between phrases which are used when the user selects other ways of reading

Some features of Japanese language are utilized in the *network structure* composition subsystem. Some examples of them are as follow.

1. In general, a Japanese phrase consists of both an independent word and dependent words. The prefix word and/or the suffix word are sometimes adjoined. The number of dependent words is not so many as compared with independent words. It seems to be efficient to analyze dependent words first. Thus, the analysis is accomplished from the end of the region to the beginning.

2. Independent words mostly include non-kana characters, alternately, dependent words are written in kana characters. Therefore, higher priority is given both to independent words which include a non-kana characters and to dependent words which consist of only kana characters.

3. The number of Kanji characters is far greater than that of kana characters. Therefore, it seems efficient to use a Kanji character as the search key to scan the dictionary indexes. These indexes are designed so that the search key must be a non-kana character in cases where there is one or more non-kana character.

4.2 Speech Information Organization Subsystem

With the user's request for speech synthesis, the speech information organization subsystem is activated. This subsystem determines the best sentence (a combination of phrases) by examining the phrases in *network structure*. After organizing the sentence, the information for speech synthesis is then organized. The pronunciation and accent of each word are determined by using the dictionaries. The structure of the sentence is obtained by analyzing the relationship between phrases. In case of numerical words, such as 1,234.56, a special procedure is activated to generate the reading. In case the user requests other ways of reading the sentence, the subsystem chooses other phrases in *network structure*, thus organizing the speech synthesis information.

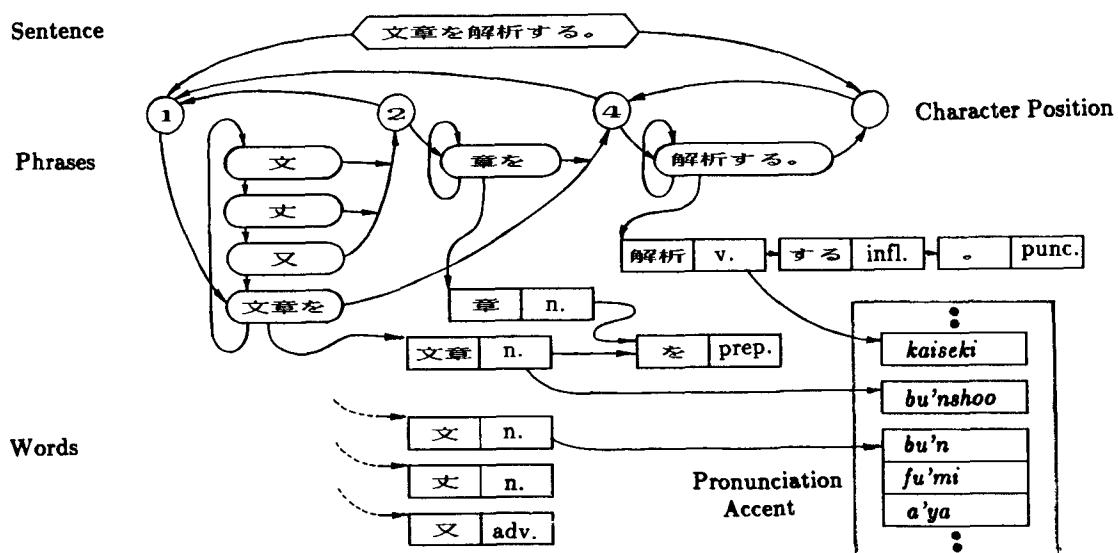


Figure 6. *Network Structure Example.*

In order to determine the most probable phrase combination in *network structure*, heuristic rules are applied. The rules have been obtained mainly by experiments. Some of them are as follow.

[1] Number of Phrases in a Sentence

The sentence which contains the least number of phrases will be given the highest priority.

[2] Probabilities of Characters

The phrase which contains more probable character candidates will be given higher priority. This probability is obtained as the result of character recognition.

[3] Written Format of Words

Independent words written in kana characters will be given lower priority.

Independent words written in one character will be also given lower priority.

[4] Syntactical Combination Appearance Frequency

The frequently used syntactical combination will be given higher priority.

(e.g. noun-preposition combination)

[5] Selected Phrases

The phrase which once has been selected by a user will be given higher priority.

In the case of Figure 3, the best way of arranging phrases is determined by applying the heuristic rule [1].

4.3 Word Dictionaries

Dictionaries used in this system are the following.

(1) Independent Word Dictionary

Nouns, Verbs, Adjectives, Adverbs,
Conjunctions etc.
65,850 words

(2) Proper Noun Word Dictionary

First Names, Last Names, City Names etc.
12,495 words

(3) Dependent Word Dictionary

Inflection Portions for Verbs and Adjectives.
They are used for conjugation.
their usage.
560 words

(4) Prefix Word Dictionary

153 words

(5) Suffix Word Dictionary

725 words

Each word stored in these dictionaries has the following information.

- (a) written mixed Kanji-kana string (first-choice)
- (b) syntactical meaning
- (c) pronunciation
- (d) accent position

Items (a) and (b) of all words are gathered to form the following four indexes.

- * Kana Independent Word Index
- * Kana Dependent Words and Kana Suffix Word Index
- * Non-Kana Word Index
- * Prefix Word Index

These indexes are used by the *network structure* composition subsystem. Items (c) and (d) are used by the speech information organization subsystem.

5. Experimental Results

Some experiments have achieved in order to evaluate the sentence analysis method. In this section, these experimental results are described.

5.1 Pronunciation Accuracy

The accuracy of pronunciation has been evaluated by counting correctly pronounced characters. In this experiment, character code strings were used as the input data. The following two whole books are analyzed.

• *Tetsugaku Annai* (Introduction to Philosophy)
by Tetsuo Tanikawa (an essay)

• *Touzoku Gaisha* (The Thief Company)
by Shin-ichi Hoshi (a collection of short stories)

As shown in Table 1, 99.1% of all characters have been given their correct pronunciation.

Table 1. Score for Correct Pronunciation.

Total Characters	128,289 (100%)
Correct Characters	127,108 (99.1%)

The major cases for mispronunciation are as follows.

- (1) Unregistered words in dictionaries
 - (1-a) uncommon words
 - (1-b) proper nouns
 - (1-c) uncommon written style
- (2) Pronunciation changes in the case of compound words
- (3) Homographs
- (4) Word segmentation ambiguities
- (5) Syntactically incorrect Japanese usage

5.2 Efficiency as the Postprocessing Roll for Character Recognition

The efficiency as the postprocessing roll for character recognition has been evaluated by comparing the characters used for speech synthesis with the character recognition result. Twelve pages of character recognition results (four pages of three books) have been analyzed. The books used as the input data are as follow.

- *Tetsugaku Annai* (Introduction to Philosophy)
by Tetsuzo Tanikawa (an essay)
- *Touzoku Gaisha* (The Thief Company)
by Shin-ichi Hoshi (a collection of short stories)
- *Yujo* (The friendship)
by Saneatsu Mushanokouji (a novel)

Table 2 shows scores for the character recognition result.

Table 2. Character Recognition Result.

Total Characters	6,793 (100%)
Correct Characters (at 1st Ranking)	6,757 (99.5%)
Correct Characters (in 1st to 5th Ranking)	6,783 (99.9%)

Table 3 shows the score for characters which are chosen as correct characters by the sentence analysis method, as well as the score for correctly pronounced characters.

Table 3. Scores after Sentence Analysis.

Total Characters	6,793 (100%)
Characters Treated as Correct Characters	6,772 (99.7%)
Characters Correctly Pronounced	6,728 (99.0%)

As shown in Tables 2 and 3, the score for correct characters obtained after the sentence analysis was 99.7%, while the score for the 1st ranking characters obtained in the character recognition result was 99.5%. This experimental result reveals that the sentence analysis method is effective as a postprocessing roll of character recognition. The state of errors found during the experiment is shown in Table 4. The difference between (b') and (b3) in Table 4 indicates the effectiveness of the sentence analysis method. The score 99.0% in Table 3 indicates the efficiency of the sentence analysis method in the book reading machine.

Table 4. State of Errors.

<< Character Recognition Error >>		
(a)	1st Ranking Chars are Incorrect	36
(a1)	Correct Chars in 2nd-5th	26
(a2)	Not among Candidates	10
<< Sentence Analysis Error >>		
(b)	Total Incorrect Char	21
(b1)	Incorrect Chars among (a1)	4
(b2)	Incorrect Chars among (a2)	10
(b3)	Incorrect Chars While Char Recognition was Correct	7
(b')	Correct Chars While the 1st Ranking Chars were Incorrect (b' = a1 - b1)	22

5.3 Efficiency of Selection by Manual

To examine the efficiency, an experiment has been conducted where sentences have been read both automatically and with the help of manual manipulation. The same text used in Section 5.2 was used in this experiment. Table 5 shows scores for the correctly pronounced characters. As shown in Table 5, 99.9% and 99.8% of all characters were given correct pronunciation after the manual selection, while 99.3% and 99.0% of all characters had been given their correct pronunciation before the manual selection, respectively. These scores reveal that most mispronunciation could be recovered by manual selection so that nearly all accurately pronounced reading can be taped.

Table 5. Scores for Characters.

Total Characters	6,793 (100%)
<< Input Data is Correct Characters >>	
Before Selection	6,745 (99.3%)
After Selection	6,787 (99.9%)
<< Input Data is Recognized Characters >>	
Before Selection	6,728 (99.0%)
After Selection	6,777 (99.8%)

6. Conclusion

A sentence analysis method used in a Japanese book reading machine has been described. Input sentences, where each character is allowed to have other candidates, are analyzed by using several word dictionaries, as well as employing syntactical examinations. After generating *network structure*, heuristic rules are applied in order to determine the most desirable sentence used for speech information generation. The results of experiments reveal: 99.1% of all characters used in two whole books have been correctly converted to their pronunciation. Even when the character recognition result is ambiguous, correct characters can often be chosen by the sentence analysis method. By manual selection, most incorrect characters can be corrected.

Currently, the authors are improving the sentence analysis method including the heuristic rules and the contents of dictionaries through book reading experiments and data examinations. This work is, needless to say, aimed in offering better quality speech to the blind users in a short computing time. Authors are expecting that their efforts will contribute to the welfare field.

ACKNOWLEDGEMENTS

The authors would like to express their appreciation to Mr. S. Hanaki for his constant encouragement and effective advice. The authors would also like to express their appreciation to Ms. A. Ohtake for her enthusiasm and cooperation throughout the research.

This research has been accomplished as the research project "Book-Reader for the Blind", which is one project of The National Research and Development Program for Medical and Welfare Apparatus, Agency of Industrial Science and Technology, Ministry of International Trade and Industry.

REFERENCES

- << in English >>
- Allen, J., ed., 1986 *From Text to Speech: The MITalk System*. Cambridge University Press.
- Allen, J. 1985 Speech Synthesis from Unrestricted Text. In Fallside, F. and Woods, W.A., eds., *Computer Speech Processing*. Prentice-Hall.
- Allen, J. 1976 Synthesis of Speech from Unrestricted Text. *Proc. IEEE*, 64.
- Allen, J. 1973 Reading Machine for the Blind: The Technical Problems and the Methods Adopted for Their Solution. *IEEE Trans.*, AU-21(3).
- Kabeya, K.; Hakoda, K.; and Ishikawa, K. 1985 A Japanese Text-To-Speech Synthesizer. *Proc. A VIOS '85*.
- Klatt, D.H. 1986 Text to Speech: Present and Future. *Proc. Speech Tech '86*.
- Klatt, D.H. 1982 The Klattalk Text-to-Speech System. *Proc. ICASSP '82*.
- Mitome, Y. and Fushikida, K. 1986 Japanese Speech Synthesis System in a Book Reader for the Blind. *Proc. ICASSP '86*.
- 1985 *Kurzweil Reading Machine Update*. Kurzweil Computer Products.
- << in Japanese >>
- Fukushima, T.; Ohyama, Y.; Ohtake, A.; Shutoh, T; and Shutoh, M. 1985 A sentence analysis method for Japanese text-to-speech conversion in the Japanese book reading machine for the blind. *WG preprint*, Inf. Process. Soc. Jpn., WGJDP 2-4.
- Mitome, Y. and Fushikida, K. 1985 Japanese Speech Synthesis by Rule using Formant-CV, Speech Compilation Method. *Trans. Committee on Speech Res.*, Acoust. Soc. Jpn., S85-31.
- Tsuji, Y. and Asai, K. 1985 Document Image Analysis, based upon Split Detection Method. *Tech. Rep.*, IECE Jpn., PRL85-17.
- Tsukumo, J. and Asai, K. 1985 Machine Printed Chinese Character Recognition by Improved Loci Features. *Tech. Rep.*, IECE Jpn., PRL85-17.