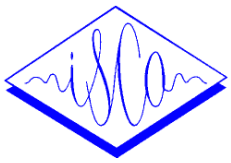

3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

Using Accent Information to Correctly Select Japanese Phrases Made of Strings of Syllables

Tetsuo ARAKI+ Satoru IKEHARA++ Hideto YOKOKAWA+

+Faculty of Engineering, Fukui University
Fukui, 910 JAPAN

++NTT Communication Science Laboratories
1-2356 Take Yokosuka-Shi Kanagawa 238-03 Japan

Abstract

This paper proposes a method to determine the most suitable string of syllable candidates using 2nd-order Markov model of accent information added to syllable characters, assuming that the correct positions of accents can be obtained from speech by acoustic processing.

From the experiment which uses the statiscal data for 5 issues of a daily Japanese newspaper, the accuracy rate that the first candidate of syllable strings are correct was shown to be improved by 6.4 %,

1.Introduction

There are two main speech recognition methods in acoustic processing. The first is the method used in isolated word recognition systems, where each isolated spoken word is idntified by DP-matching with stored standard syllable string patterns [1]. This method has a high recognition rate, but the number of recognizable words(i.e., the vocabulary) is generally limited to several hundred. Therefore, this method is not suitable for spoken word recognition for a large vocabulary.

The second is the method to partition continuous speech into segments and to classify each segment according to a set of phonemes or syllables [2] [3] [4]. This method is thought to be suitable for speech recognition of all spoken(i.e.,natural) Japanese sentences, however several problems in mis-segmentation (which leads to insertion and deletion errors) and misclassification (which leads to substitution errors) must be solved. In the speech recognition systems, insertion and deletion probabilities are generally quite small compared to substitution probabilities. Therefore, only substitution errors are discussed in this paper, assuming that the segmenting process is perfect(that is, insertion and deletion errors are excluded).

In speech recognition systems for naturally spoken Japanese sentences, phonological, morphological, syntactic, and semantic properties are generally considered useful in reducing ambiguities (sustitution, insertion and deletion errors). Natural language processing has been applied to the syllable candidate lattice obtained by the recognition algorithm based on acoustic characteristics. Up to now, 2nd-order Markov model of syllable characters is known to be useful to solve this ambiguity of syllable candidates [7] [8].

This paper proposes a method to determine the most suitable string of syllable candidates using 2nd-order Markov model of accent information added to syllable characters, assuming that the correct positions of accents can be obtained from speech by acoustic processing. The effectiveness of this method is evaluated by the experiment which uses the statiscal data for 5 issues of a daily Japanese newspaper.

2. Basic Definitions and the method to determine the correct string of syllable candidates

The Japanese speech input system model shown in Fig.1, is used to evaluate the effectiveness of the method to determine the correct string of syllable cadidates (called syllable lattce) output from speech recogvnition device (accoustic processing) [2] using Markov models.

Japanese sentence can be separated into syntactic units called "bunsetsu", where a "bunsetsu" is composed of one *independent word* and a sequence of n (equal to or more than 0) *dependent words*. Thus, "bunsetsu = (*independen word*)(*dependent word*)^n, where *independent word* = noun, verb, adverb, adjective, verbal adjective, attributive, conjunction, or interjection,etc., and *dependent word* = auxiliary verb or post positional word.

Furthermore, it is known that strings of phoneme and informations of accent and pause can be obtained through a Japanese text to speech system [6] for Japanese "kanji-kana" sentences input to it .

First, two types of Markov models for syllable characters are defined.

(Definition 1) A syllable string is deboted by X=$s_1 s_2 \cdots s_n$. When the accent is added to the syllable s_i in the syllable string X , the syllable s_i is called *syllable with accent* (denoted by $\hat{s}$), and when the accent is not added to syllable s_j, s_j denotes *bare syllable*.

Furthermore when all syllables in a syllable string X are *bare syllables*, X is called a *syllable string whitout accent* and when *syllables with accent* are included in a string X, X is called a *syllable string with accent*. Then Markov chain probability obtained using *syllable strings without accent* are called *Markov probability WithOut Accent* (denotes *MWOA*)), and

that of *syllable strings with accent* are called *Markov probability With Accent* (denotes MWA). ∎

In this paper 2nd-order Markov chain probability are used.

Next, two types of syllable lattices with accent are presented:

(Definition 2) A syllable lattice L is called a *Syllable Lattice with Accent located Correctly* (denotes $SLAC$) when the location of accent are correctly indicated in L. L is also called *Syllable Lattice with Accent located Ambiously* (denotes $SLAA$) when the location of accent are not correctly indicated in L. Here it is assumed that accent for syllable string obtained from L are all located at the same position. ∎

An example of $SLAC$ is shown in Fig.3.

Finally, the methods to determine the correct string of syllable candidates from a syllable lattice using two types of Markov models described abve are presented as follows:

(Definition 3) The most suitable string of syllable candidates obtained from a syllable lattice $L=s_1 s_2 \cdots s_n$ is defined to be a string which has the smallest value of folowing equation computed using 2nd-order Markov probanbility corresponding to each string:

$$-\sum_{i=1}^{n+} \log_2 p(x_i | x_{i-2} x_{i-1})$$

Here x_j denotes a space symbol b if $j < 0$ *or* $j > n$.

The method to determine the correct string of syllable candidates obtained from $SLCA$ and $SLAA$ using MWA is called Method by MWA (denotes $MMWA$), and that of $MWOA$ is called Method by $MWOA$ (denotes $MMWOA$). ∎

Each column of the syllable lattice consists of ten candidates and a correct syllable is included in these candidates. In our experiment, 110 syllables are used for the complete set of Japanese syllables, including nasal sounds, long vowels and contacted sounds. Each syllable lattice is produced by a confusion matrix(a list of possible substitution errors for each syllable) [5] based on the characteristics of syllable recognition candidates output from typical speech recognition devices with the following restrictions. (1) Substitution errors can arise only among syllables which have the same vowel as the correct syllable (e.g., "a" can substitute for "ta" or "pa" but not to "i" or "ku" etc.), (2) A double consonant, "n" and a log vowel in Japanese are never substitution errors. Also substitution errors involving nasal sounds are the same as for their repective "k" sounds ("ka", "ki", "ku", "ke" and "ko"), (3) Average accuracy syllable selection from candidates is abut 89-percent.

Especially, the segmentation process in acoustic processing is assumed to be perfect (insertion or deletion errors), and each syllable lattice is assumed to always contain the correct sylable string. An example of syllable lattice is shown in Fig.2. From this fugure, it is seen that a lot of syllable string candidates may be derived from a syllable lattice. (e.g., "ni-hon-wa", "i-kon-wa" or "ni-ho-na" etc.)

3. Experimental Results

3.1 Experimental Conditions

(1) The type of sentences: Japanese newspaper articles.
(2) The number of "bunsetstu": 52,642 "bunsetstu"
(3) The total number of syllables:353,065 syllables
(4) The type of 2nd-order Markov probability: $MWOA$ and MWA
(5) The type of 2nd-order Markov probability: $MWOA$ and MWA
(6) The type of input phrases: "bunsetsu" strings and strings partitioned by pauses.

3.2 Experimental Results

The accuracy ratio of determining correct string of syllable candidates obtained from $SYAA$ and $SYAC$ using MWA and $MWOA$ are obtained in the following two casees:

(i)case of "bunsetsu" syllable strings for *closed data* (shwon in Fig.4) and *open data* (shown in Fig.5)
(ii) case of syllable strings partitioned by pauses for *closed data* (shown in Fig.6) and *open data* (shown in Fig.7).

(1) The entropy of $MWOA$ and MWA

2nd-order Markov probability are obtained using the statistical data for 5 issues of a daily Japanese newspaper. The saturation curves of 2nd-order Markov probability using statistical data of syllables are shown in Fig.2. From this figure, it is seen that 80 % of rules, which denote the set of three tupples (x_{i_2}, x_{i-1}, x_i) that 2nd-order Markov probability $\mathrm{p}(x_i \mid x_{i-1} x_{i-2})$ is not equql to 0, are saturated. The percentage of rules saturated improves as the statistical data increases, the saturation degree of 2nd-Markov probability. The sdmaller the entropy of Markov probability is, the more effective the decision of correct candidates using Markov probability is expected. The entropy of 2nd-order Markov probability MWA and that of $MWOA$ are 2.96 and 2.98 respectively.

(2)The effect of the methods $MMWA$ and $MMWOA$ to determine correct syllable strings

Compared to the method $MMWOA$, the accuracy rate that the first candidate of syllable stings are correct was improved by 6.4 % by the method $MMWA$. The accumulative accuracy rate for first ten candidates in case of $MMWOA$ can be obtained by that of first four candidates in case of $MMWA$. Especially the method $MMWA$ is effective to determine the correct syllable strings for *closed data*, but the value of accuracy rate obtained using the method $MMWA$ applied to *open data* can be also expected to be improved if the statistical data for 30 issues of a daily Japanese newspaper obtains.

From these results, it is seen that the method $MMWA$ is effective to determine the correct syllable strings.

(3)The Effect of $MMWA$ applied to the syllable strings partitioned by pause information

From Fig.5 and Fig.7, it is seen that the 1st accuracy rate for the strings partitioned by pause is inferior to that of "bunsetsu" strings, but the 10th accumulative accuracy rate for the strings partioned by pause approach to that of "bunsetsu" strings.

The examples of correct or wrong (means that the first candidats are correct or wrong) "bunsetsu" strings determined by *MMWA* and *MMWOA* are shown in Table 1.

5. Conclusion

This paper proposes a method to determine the most suitable string of syllable candidates using 2nd-order Markov model for accent information added to syllable characters, assuming that the correct positions of accents can be obtained from speech by acoustic processing.

From the experiment which uses the statiscal data for 5 issues of a daily Japanese newspaper, the value of accuracy rate that the first candidate of syllable stings are correct was shown to be improded by 6.4 %.

It is a further subject to develop the method to determine the correct string of syllable candidates obtained from the actual syllable lattices, which include the syllables substituted, inserted and deleted wrongly, using accent information.

References

[1] H.Sakoe and S. Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition", *IEEE Trans. Acoust., Speech, Signal Processing* , ASSP-26, 1, pp43-49 (1978)

[2] R.Nakatsu and M.Koda,"An Acoustic Processor in a Conversational Speech Recognition System", *Trans. IECE Japan*,Vol.J61-D,No.4, pp261-268 (1978)

[3] B.Aldefeld, L.R. Rabiner, A.E. Rosenberg and J.E.Wilpon, "Automated Directory Listing Retrieval System Based on Isolated Word Recognition", *Proc. IEEEE* , Vol.68, No.11, pp.1364-1379 (1980)

[4] L.R.Rabiner,S.E.Levinson and M.M. Sondai, "On the Application of Vector Quantization and Hidden Markov Models to Speaker-independent, Isolated Word Recognition", *Bell System Technical Journal*, Vol.62, No.4, pp.1075-1105 (1983)

[5] H.Hamada, I.Namiki and R.Nakatsu,"Japanese Sentence Input System by Character Unit Speech", Trans Committee on Speech Research Acoust. Soc. Jap., S84-24 (1984)

[6] Y. Ooyama and Y. Miyazaki, "Natural Language Processing in a Japanese-text-to-speech System ", *Information Processing Society of Japan*, Vol.27, No.11, pp.1053-1061 (1986)

[7] T.Araki,J.Murakami and S.Ikehara, "Effect of Reducing Ambiguity of Recognition Candidates in Japanese Bunsetsu Unit by 2nd-Order Markov Model of Syllables", *Information Processing Society of Japan*, Vol.30, No.4, pp.467-477 (1989)

[8] J.Murakami,T.Araki and S.Ikehara, "The Effect of Trigram Model in Japanese Speech Recognition", *The Institute of Electronics, Information and Communication Engineers*, Vol.J75-D-II, No.1, pp.11-20 (1992)

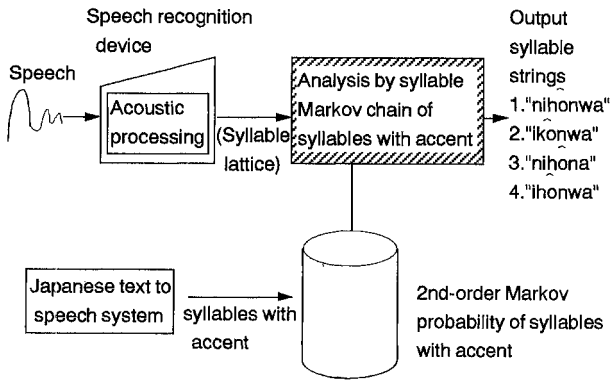

Fig.1 A Model of Japanese Speech Input System

MMWOA	Correct Syllable String	nihonwa、sanjyuu、sekiyukikio、kaigaidewa
	Wrong Syllable String	ima、sengo、keizaikikio、norikoe、baneni 、sugata、sanaka、wasureruna
MMWA	Correct Syllable String	îma、nihônwa、sânjyuu、sekiyukîkio 、keizaikikio、norikoe、sugata、sânaka、wasureruna
	Wrong Syllable String	sêngo、bâneni、kâigaideha

Table.1 Example of Correct and Wrong Strings of Syllable Obtained by MMWA and MMWOA

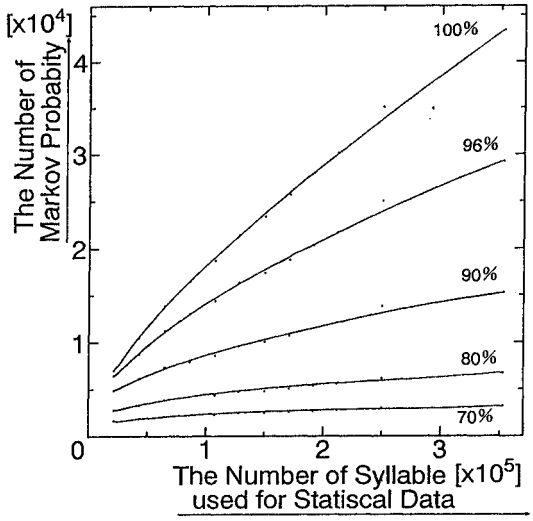

Fig2. Learning Characteristics of MWA

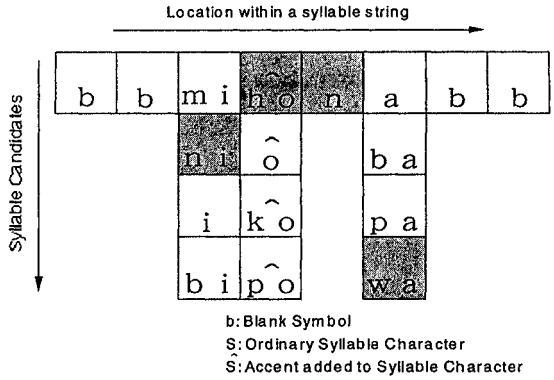

Fig.3 An Example of SLCA

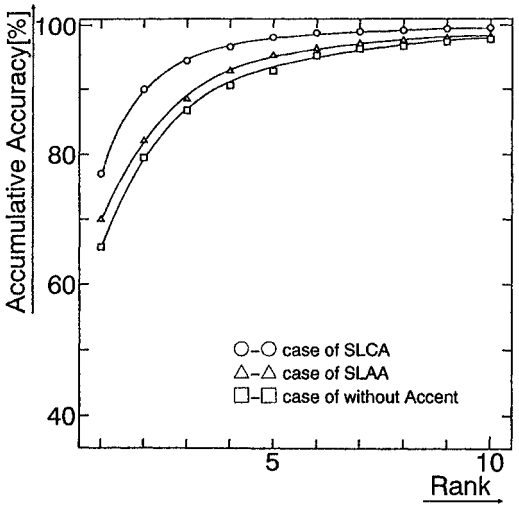

Fig.4 Experimental Results for "Bunsetsu" Syllable Lattices(Close Data)

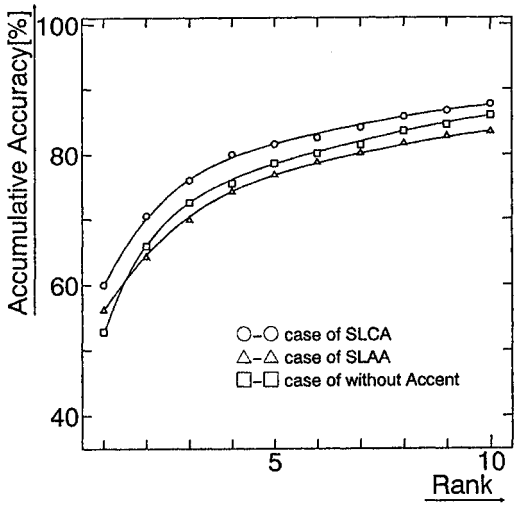

Fig.5 Experimental Results for "Bunsetsu" Syllable Lattices(Open Data)

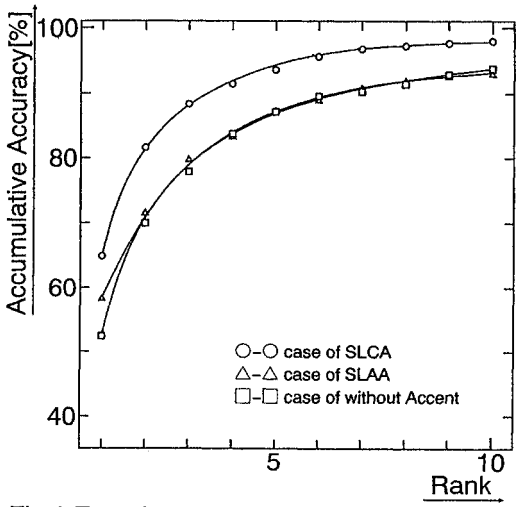

Fig.6 Experimental Results for Syllable Lattice Partioned by Pauses(Open Data)

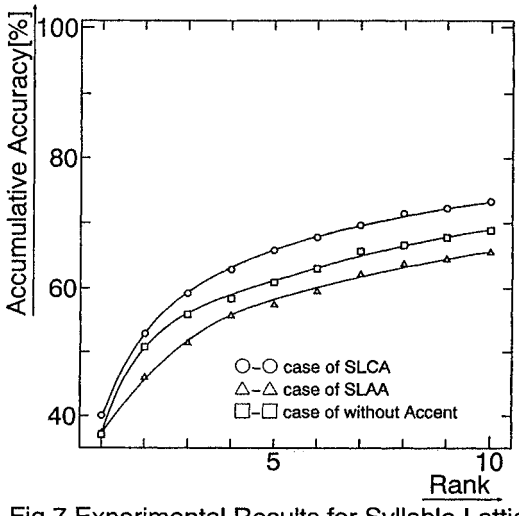

Fig.7 Experimental Results for Syllable Lattice Partioned by Pauses(Open Data)