TR-I-0030

Analysis and Prediction of Vowel-Devocalization in Isolated Japanese Words

日本語孤立発声単語における母音無声化の分析と予測

Hisao Kuwabara and Kazuya Takeda
桑原尚夫、武田一哉

June, 1988

Abstract: An analysis of vowel-devocalization in Japanese is presented as a method of finding general rules for allophonic variations of phonemes in continuous speech, i.e. to find quantitative relationships between phoneme environments and allophonic variations. A speech database consisting of about 5,000 common Japanese words spoken in isolation by a professional male announcer was used. The database used is part of a large-scale speech database that is now being compiled and hand-transcribed at several levels of phonetic detail. Based on the acoustic-phonetic transcription of the speech data, factor analysis of the phonetic environment was performed for the devocalized vowels. The results reveal that the influence of the immediately following phoneme on devocalization is the greatest of the four factors examined. The second most important factor is the position of the accent nucleus. Automatic prediction of vowel devocalization is discussed using the results of this analysis.

1. Introduction

Current continuous speech recognition and synthesis-by-rule systems generally require the knowledge of acoustic variations of phonemes in continuous speech caused by coarticulation. However, it is almost impossible to determine the general rules for these phonetic variations. In this study, therefore, attention was focused on one aspect of allophonic variations: a phenomenon of vowel devocalization (or equivarently syllable devocalization).

Using a large-scale speech database now under construction at ATR, devocalized vowels have been analyzed to find the relationship between the phenomenon and the phonetic environments of the vowels. Several phonetic environments that would cause vowel devocalization are taken into consideration and a prediction method for devocalization has been proposed through a factor analysis based on the second method of quantification.

2. Speech Database

A large scale speech database was used to analyse devocalized vowels. The database consists of approximately 5,000 common Japanese words, 200 phonetically balanced words, 100 Japanese CV-syllables, and small number of short sentences. These words were spoken in isolation by a professional male announcer and the speech data were mannualy labelled with phonetically fine transcription by inspecting the spectrograms.(1) A database management system based on UNIFY, a database management tool, access easily to both speech and label files has been established.(2)

Fig. 1 represents an example of spectrogram of the word "a to shi ma tsu" and its acoustic/phonetic transcriptions. Multiple transcriptions were made in five different ways: 1) phonetic symbols, 2) acoustic events where fine acoustic-phonetic symbols are given to describe the acoustic variations of the utterance, 3) alophonic variants, 4) inseparable portions, and 5) vowel centers.

Vowel devocalization is indicated on the 'alophonic variants' layer in the database by the symbol "dv" as shown in Fig. 1. The criterion for deciding whether a vowel is devocalized or not is the presence or absence of the voice-bar (low frequency energy) or the periodicity observed on the spectrogram. There are a total of 534 devocalized vowels in the database examined. Details are noted in Table 1 in combination with consonants.

3. Devocalization and Phonetic Environments

Among the five Japanese vowels, only two vowels, /i/ and /u/, are found to devocalize depending on their phonetic environments.(3) The remaining three vowels seldom devocalize. According to Table 1, the largest number of vowel devocalizations is observed for vowel /i/, especially when it is connected with the consonant /sh/. In other words, the CV-syllable /shi/ (which is pronounced [ʃi]) is the one most likely to devocalize in various phonetic contexts. For the syllable /shi/, phonetic environments that bring about devocalization are summarized in Table 2. Table 2 (a) shows the number of devocalized /shi/ syllables that occur at three different positions in a word, and also shows those numbers in relation to the position of the accent nucleus.

As for the syllable position in a word, almost all cases of devocalization occur in syllables at the word-initial or the middle positions.(4) Only one case of devocalization is observed for a syllable in the word-final position. The number of devocalized syllables at the word-initial position is greater than those at the middle position. As for the influence of the accent nucleus, it is obvious that the syllables do not devocalize at all when they are accentuated. However, a relatively high devocalization rate is found when the immediately preceding syllables are accentuated.

Table 2 (b) shows the numbers and rates of devocalized /shi/ syllables against immediately following consonants. It is noted that all /shi/ syllables are devocalized when followed by the consonant /k/, and 94 percent when followed by the consonant /t/. Fifty percent of the syllables are devocalized when the voiced consonant /g/ follows, and relatively small number of syllables are devocalized when the nasal consonants follow.

Next, our results are compared with the entry in the accent-dictionary compiled by NHK.(5) According to the dictionary, trends for vowel-devocalization are described using three kinds of phonetic environments that greatly influence the phenomenon. They are 1) the immediately following consonant, 2) the position of accentuation, and 3) the position of breath break (pause). Two devocalization rules are derived from the three environments. Rule 1 is the description in relation to following consonants. Rule 2 is for accentuation and pause. Fig. 2 summarises our results for /shi/ syllables according to these two rules. The /shi/ syllables examined fall into either a devocalized group (Yes) or a non-devocalized group (No) according to each rule, and the numbers in the table show speech data which conform to the rules. For example, among 116 syllables which fall into the category 'Yes' of Rule 1 (which means 116 syllables will devocalize according to this rule), 104 actually devocalized. The remaining 12 did not. The discrepancy for Rule 2 is very large because only one sample is found which actually devocalized among 29 samples for which devocalization is expected according to Rule 2. This will probably come from the speech data analysed. NHK's rules are

derived from continuous speech data while results of this study are obtained from the isolated words.

4. Factor Analysis and Prediction of Devocalization

As observed in Section 3, frequency of devocalization depends on various phonetic environments. A quantitative analysis will be necessary to evaluate how each of these environmental factors affects vowel devocalization. In this section, results of the factor analysis on devocalization using the second theory of quantification are presented. This theory is an analysis method, assuming several categories as the factors, that gives each category a numerical value so that the influences from the categories are well balanced in order to reflect the real data. Therefore, the larger the numerical value of a factor, the greater the influence from the category to which the factor belongs. The prediction of devocalization is possible since a linear combination of the factors forms a kind of decision surface as to whether or not the syllables in question would devocalize.

Fig. 3 represents a result of the factor analysis for the devocalized syllables /fu/. First, four categories considered of great influence on devocalization are assumed: 1) the immediately preceding syllable, 2) the immediately following syllable, 3) position of the accent nucleus, and 4) word length in mora. For each category, a few sub-categories which are also believed influential are taken into consideration. These sub-categories are defined as the environmental factors in the factor analysis.

As a result of the analysis, a positive value assigned to a factor of a sub-category as a score stands for the negative influence on the devocalization, i.e. the syllable in question is less likely to devocalize in that phonetic environment. Conversely, if the value is negative, the syllable will be more likely to devocalize. Zero means no effect. The 'range' for each category, which is defined as the difference between the maximum and the minimum scores, represents the dynamic range within the category. The greater the range, the more influence the factor has on devocalization.

The result reveals that the immediately succeeding syllable has the greatest influence on devocalization. The second greatest factor is the accent nucleus followed by the factor of word length. The immediately preceding syllable has the least influence of all. In other words, the immediately succeeding syllable plays the decisive role in devocalization. The syllable /fu/ in particular is the most likely to devocalize when it is followed by a voiceless plosive or a voiceless fricative.

An attempt to predict the occurence of syllable devocalization was made using the result of factor analysis. Fig. 4 shows the frequency distribution for the devocalized and non-devocalized /fu/ syllables as a function score. For each /fu/ syllable, scores given to the individual factors in Fig. 3 are summed up. The summation, of course, depends on the phonetic environment of the syllable to be examined. Using the word /ka i fu ku/ (回復、'recovery' in English) as an example, the phonetic environment of the third syllable /fu/ is such a case: the preceding syllable is a vowel /i/ (factor f_2), the succeeding consonant is a voiceless stop /k/ (f_5), the vowel /i/ is accented (f_9), and the word consists of 4 morae (f_{12}). A total score of the environmental factors for this syllable will be, $f_2+f_5+f_9+f_{12}=-0.98$. The word /fu ne/ (船、'ship') is another example. In this case, the syllable /fu/ itself will be accentuated. The total score for the /fu/ in this context will be, $f_1+f_7+f_8+f_{11}=1.32$. As far as the speech database used is concerned, the /fu/ syllables which corresponded to the first example were actually devocalized and those which corresponded to the second were not.

The total score of the environmental factors for each /fu/ syllable in the database was calculated as described above, and two histograms of occurence vs. score were made for devocalized and non-devocalized syllables. Fig. 4 represents the best-fit Gaussian curves for the histograms. It is observed that the devocalized and the non-devocalized syllables seem to be well separated, and the distribution for the devocalized syllables is considerably narrower than that for the non-devocalized syllables. The result indicates that prediction as to whether or not a syllable in a word will devocalize is possible to some extent using the environmental factors shown in Fig. 3. In fact, it was found that the prediction with an accuracy of 79 percent would be possible when the vertical dashed line in Fig. 4 was used as the boundary between the two groups.

5. Summary

An analysis has been made for the vowel devocalization phenomenon as a basis for establishing quantitative rules of the acoustic variations in reference to the phonetic environments for various phonemes. In Japanese, the devocalizations are generally observed for two vowels, /i/ and /u/, in combination with five consonants /t, k, h, s, sh/ depending on their phonetic environments.

It was revealed that the most important factor which affected the devocalization as a phonetic environment was the immediately succeeding syllable. The next most important factor was the position of the accent nucleus. A vowel is most likely to devocalize when it is followed by a voiceless plausive-consonant.

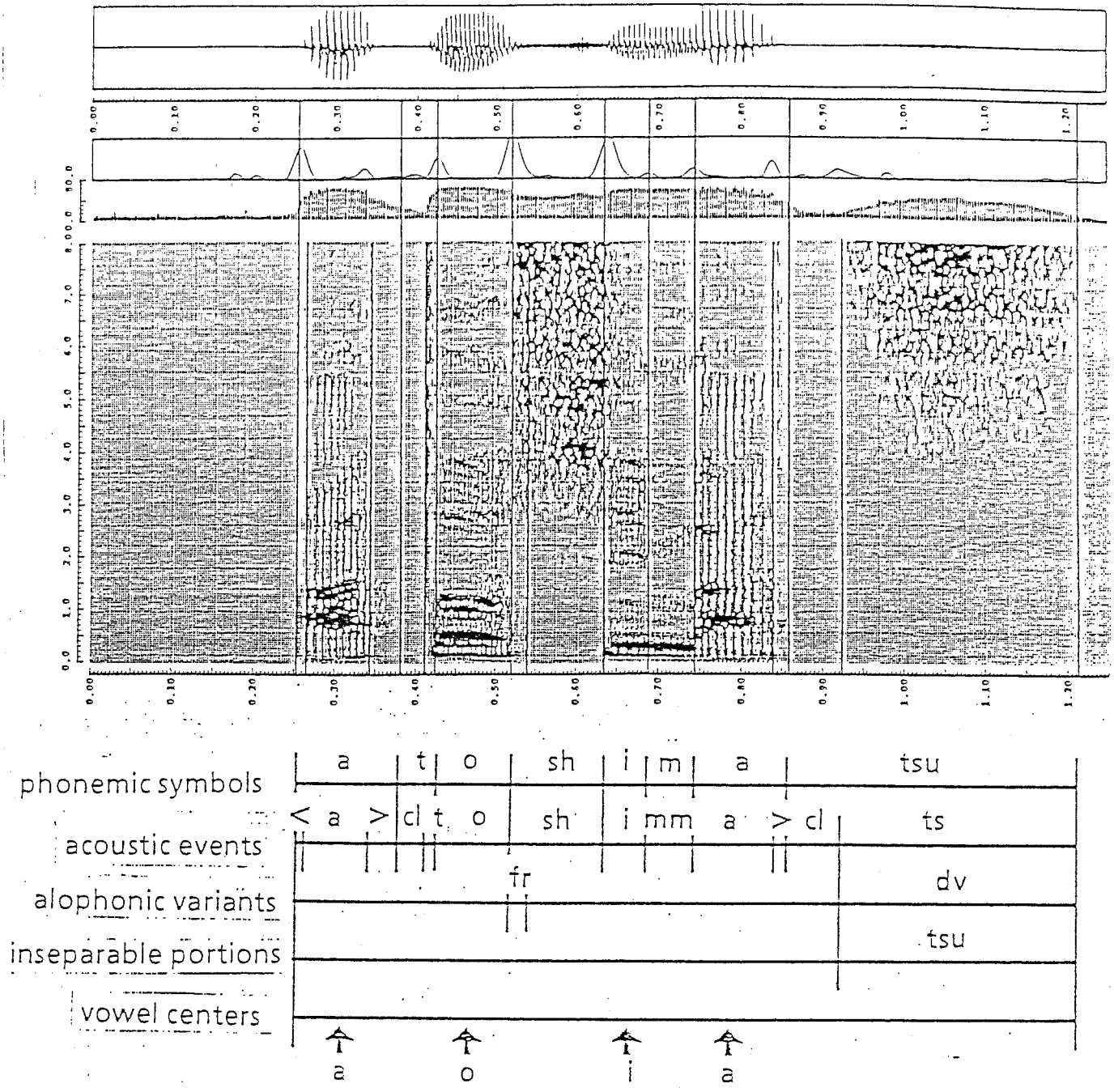
Based on the distribution, the devocalization for the syllable /fu/ was found to be predictable with an accuracy of 79 percent using the scores of environmental factors that were obtained from a factor analysis.

Acknowledgments

The authors wish to express their gratitude to Dr. A. Kurematsu, the president of ATR Interpreting Telephony Res. Labs. and to Dr. K. Shikano, the head of the Speech Processing Deptartment for their discussions and encouragement.

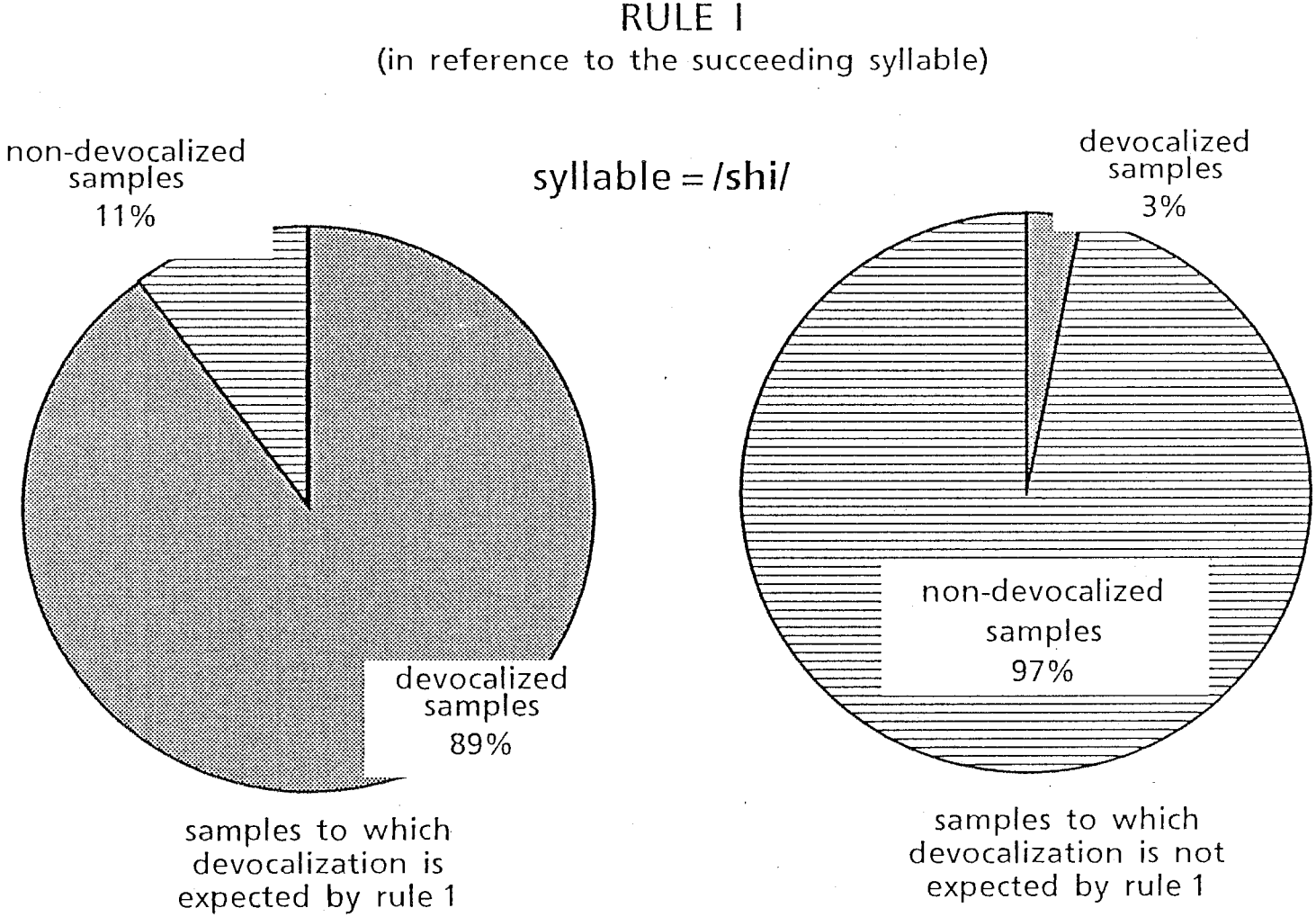
References

(1) K. Takeda, Y. Sagisaka, S. Katagiri: "Acoustic-phonetic labeling in a Japanese speech database," Meeting of Acoustical Society of Japan, 2-5-10, pp.69-70, March, 1987.
(2) K. Takeda, S. Morikawa, H. Kuwabara: "Construction of Japanese speech database management system," Meeting of Information Processing Society of Japan, 2H-4, September, 1987.
(3) K. Takeda, H. Kuwabara: "Analysis and prediction of devocalizing phenomena," Meeting of Acoustical Society of Japan, 3-3-8, pp.105-106, October, 1987.
(4) O. Kimura, N. Kaiki, A. Kito: "A study on devoicing rule in speech synthesis by rule," Meeting of Acoustical Society of Japan, 1-1-11, pp.137-138, March, 1988.
(5) NHK Accent Dictionary, revised version, Japan Broadcasting Corporation (NHK), 1985.



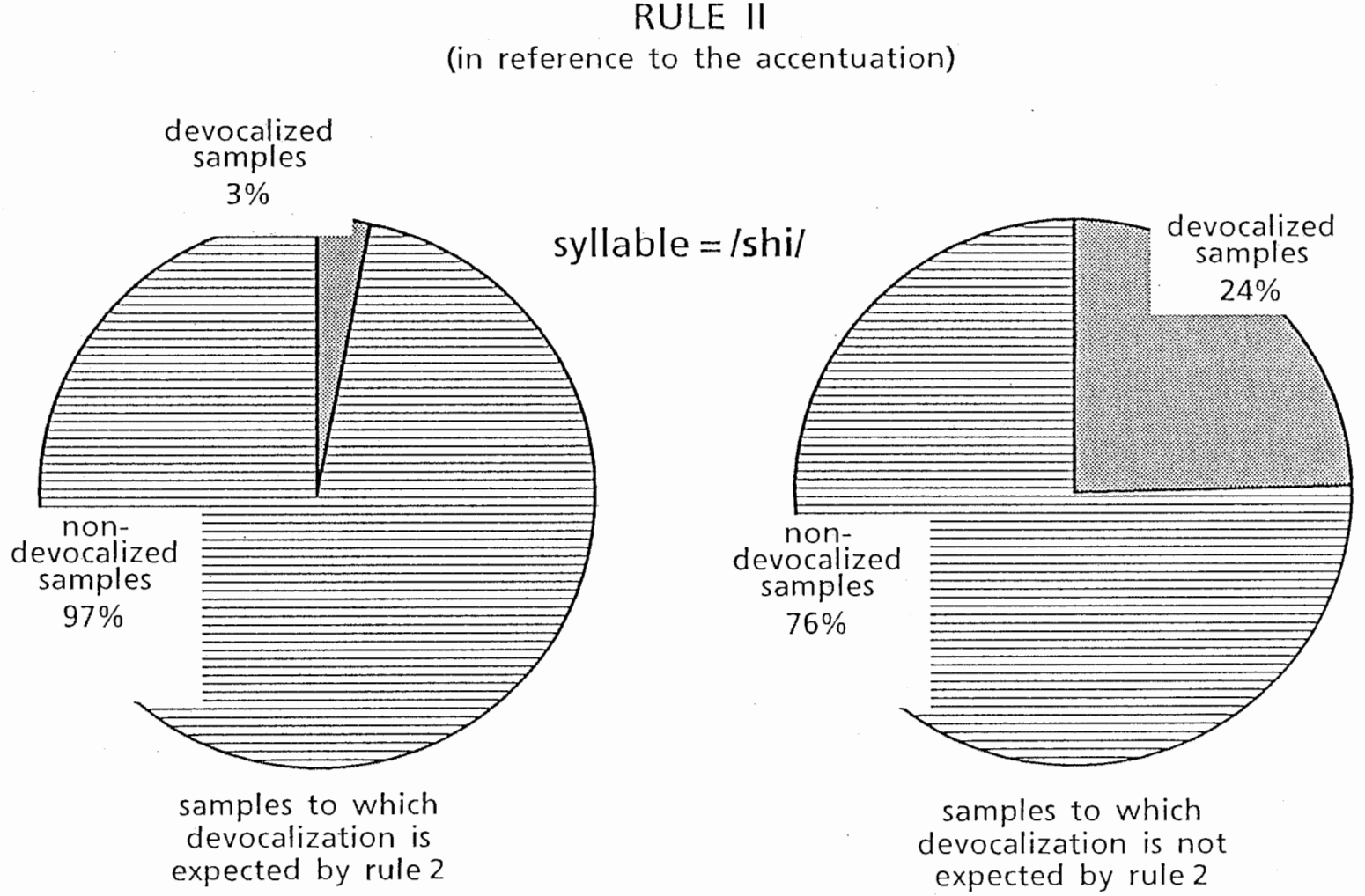
An example of multiple acoustic-phonetic transcriptions of the utterance /atoshimatsu/.

Fig. 1



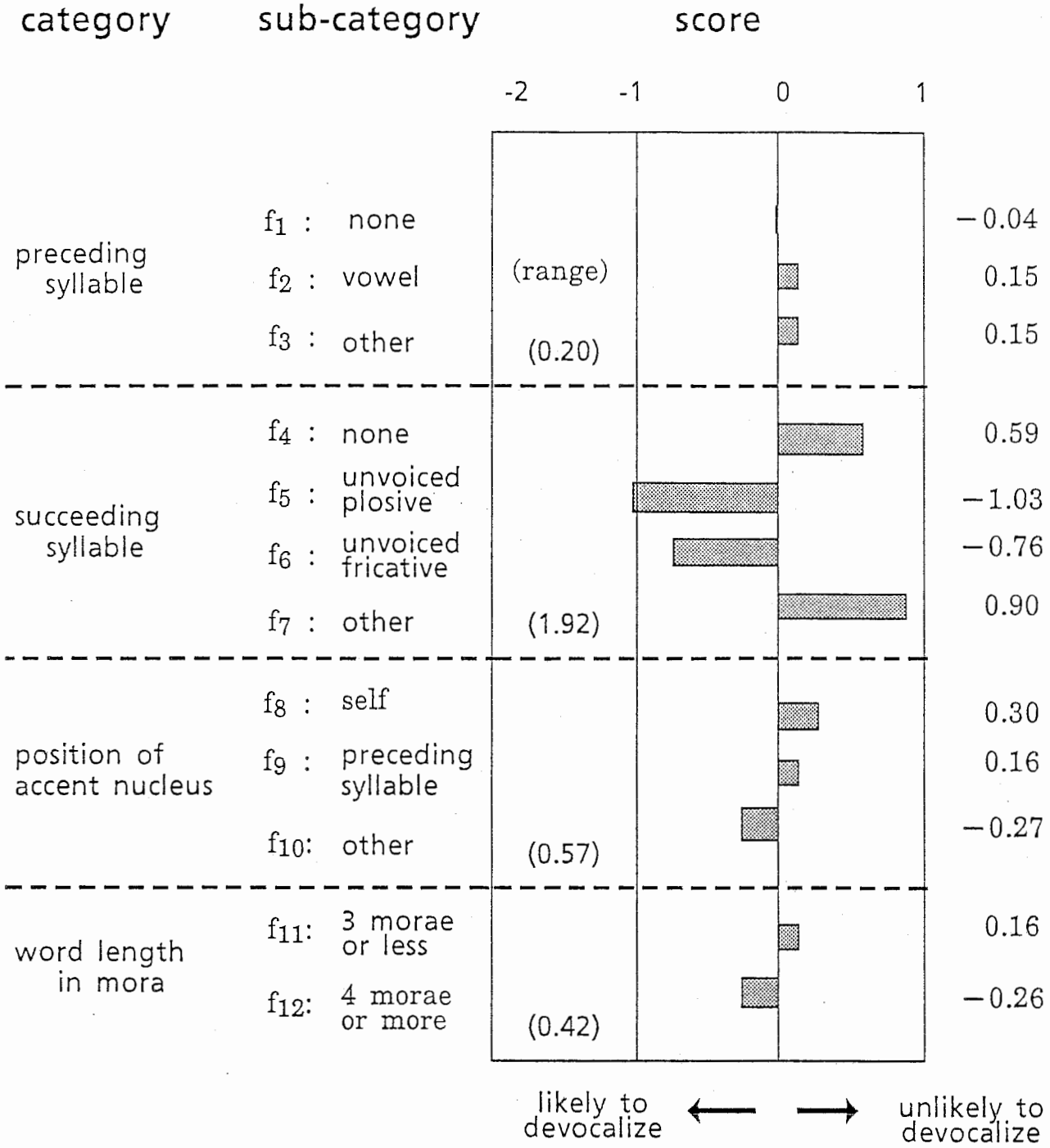
Comparison with NHK Accent Dictionary

Fig. 2 (a)



Comparison with NHK Accent Dictionary

Fig. 2 (b)



syllable = /fu/

Result of Factor Analysis

Fig. 3

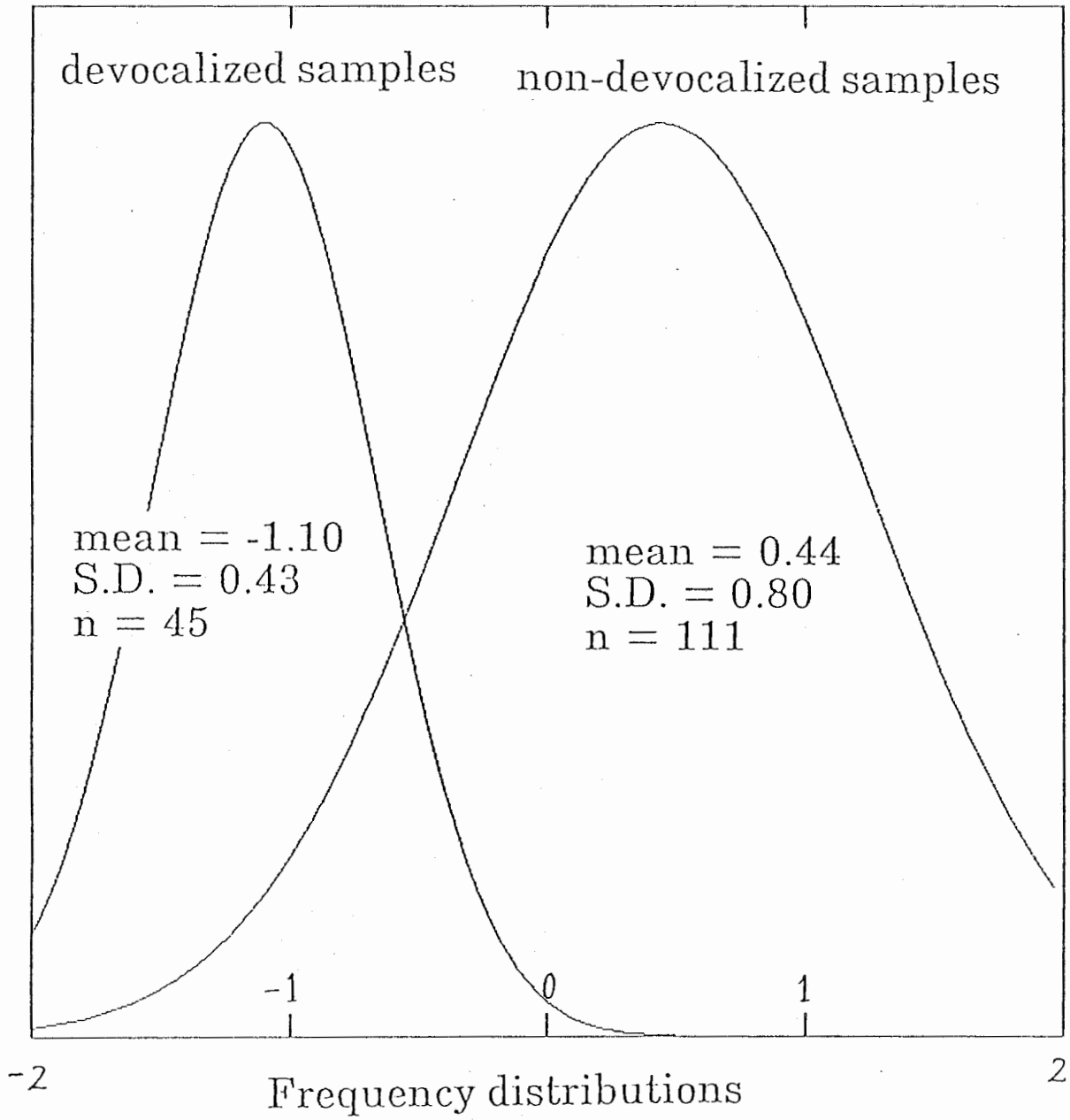

devocalized samples
non-devocalized samples
mean = -1.10
S.D. = 0.43
n = 45
mean = 0.44
S.D. = 0.80
n = 111
-1
0
1
-2
2
Frequency distributions

Fig. 4

Number of Devocalized Syllables

	k	t	s	sh	h
i	54 / 429 (13%)	28 / 200 (/chi/) (14%)		116 /499 (23%)	56 /138 (41%)
u	77 / 774 (10%)	90 / 499 (/tsu/) (18%)	41 / 366 (11%)	27 / 127 (21%)	45 / 156 (/fu/) (29%)

/ki/, /chi/, /shi/, /hi/

/ku/, /tsu/, /su/, /shu/, /fu/

Table 1

Devocalization of /shi/ Syllables vs Environmental Factors

	syllable position in a word			position of accent nucleus		
	word-initial	middle	word-final	itself	preceding syllable	other
number of devocalized syllables	70 (31%)	45 (24%)	1 (1%)	0 (0%)	53 (34%)	63 (28%)
total number	224	188	82	121	155	223

Table 2 (a)

Devocalization of /shi/ Syllables vs the succeeding consonant

succeeding consonant	k	t	g	s	n	m
number of devocalized syllables	48	56	5	3	2	1
rate of devocalized syllables	100%	94%	50%	25%	15%	3%

Table 2 (b)