

## FACTORS CONTRIBUTING TO THE DIFFICULTY IN PRONUNCIATION\*

S. Kiritani, T. Hayata\*\*, S. Okamoto, and O. Fujimura

The relation between the subjective difficulty in pronunciation of commonly used Japanese words and their phonetic features was examined.

Personal inquiries were given to one thousand radio/TV announcers for two thousand items of words and short phrases. For each item, the subjects were instructed to select one of three marks, 0, 1, and 2 according to their subjective judgements on difficulty in pronunciation. These values were averaged for all the subjects for each item  $i$ . The average value will be denoted here as  $X^i$ .

For analysis of the response, the following method was employed. The items were transcribed by using the following phonemic symbols.

Consonant : ' h p t k b d g m n ŋ s z c r

Vowel : a i u e o

Semivowel : y w

: N Q

Each syllable was assumed to contain some of the following phonetic features that are assumed to be relevant to the difficulty of pronunciation.

1. voiced
2. unvoiced
3. strident
4. interrupted
5. flap
6. nasalized
7. labial
8. dental
9. velar
10. diffuse
11. palatalized
12. labialized
13. nasal syllable lengthening /N/
14. geminate consonant /Q/

As the first approximation, each phonetic feature is considered to have certain characteristic contribution to the subjective difficulty in pronunciation. It was assumed that  $X^i$  can be approximated by  $\sum_j n_j^i x_j$ , where  $n_j^i$  is the number of occurrence of the phonetic feature  $j$  in the item  $i$ , and  $x_j$  is the magnitude of contribution of the phonetic feature  $j$ . The values of  $x_j$ 's were determined so

---

\* Work carried out in cooperation with the Japan Broadcasting Corp.

\*\* Radio & Television Culture Research Institute, Nippon Hoso Kyokai (Japan Broadcasting Corp.)

as to minimize  $\sum_i (X^i - \sum_j n_j^i \cdot x_j)^2$ . Discrepancy between the predicted value and the subjective evaluation was examined and the specific combinations of the phonetic features that may contribute to subjective difficulty in pronunciation were found. A new set of possible factors was chosen and the magnitude of contribution of each factor  $x_j$  was re-calculated in the same way as stated above. Satisfactory comparison between the predicted and observed values of difficulty has been obtained by the successive approximation.

Table 1 shows a typical result of calculation for the items of 3- and 4-syllable words.

The factors 1-6 are the individual phonetic features concerning the manner of articulation of consonants. They do not show significant contribution.

The factors 7-10 have most significant contribution. It seemed that the item in which these factors appeared in a successive position had markedly large value of  $X^i$ . In this case, the contribution of the factor  $j$  was assumed to be  $2x_j + y_j$ . The calculated values of  $y_j$ 's are given in parentheses following the value of  $x_j$ s.

The factors 7 and 8 are the diffuse vowels /i/ and /u/ placed between strident consonants or voiceless consonants. The factor 9 is the appearance of consonants characterized by the same place of articulation in the neighbouring syllables. For this factor to be counted as positive, it is necessary that the consonants in three consecutive syllables have the same place of articulation.

Concerning these three factors, existence of the phoneme /N/ or /Q/, a diphthong or an elongated vowel seemed to break the specified combination of phonetic features. Thus, the item of the CVNVCV structure, where C is a voiceless consonant and V is a diffuse vowel, is not considered to contain the factor 8. Similarly, in the case of a CVCVNVCV structure, the correction  $y_9$  is not taken into consideration.

The factor 10 is the semivowel /y/ and the factor 11 is the correction to the factor 10 when /y/ is followed by a diphthong or an elongated vowel. The factor 7 refers to the glottal phoneme that is not at the word initial position nor between the same vowels. The glottal phonemes /' / and /h/ may have a significant contribution in some specific contexts, although the exact condition is not clear as yet.

Table 1. Typical factors and their magnitudes of the contribution (see text).

Factor	$x_j \times 10^2$	
	(3 syllable)	(4 syllable)
1. voiced	+1	+2
2. voiceless	+3	+3
3. interrupted	0	-1
4. strident	+2	+1
5. flap	+2	-1
6. nasal	+3	+4
7. diffuse vowel between strident consonants	+7 (+2)	+10 (-2)
8. diffuse vowel between voiceless consonants	+9 (+5)	+2 (+4)
9. repetition of the same place of articulation	+4 (+10)	+2 (+8)
10. semivowel /y/	+12 (+10)	+8 (+10)
11. /y/, followed by diphthong or elongated vowel	-8	-8
12. glottal phoneme	+6	-1
13. glottal phoneme between the same vowel	+11	+9
14. /N/	-5	-7
15. /Q/	+5	-2
16. diphthong, elongated vowel	-3	-3