SOME CHARACTERISTICS OF PITCH VARIATIONS IN JAPANESE

Summary

An investigation of sentence intonation of Japanese on the basis of Kymograph records of the speech of two subjects who speak the Tokyo dialect demonstrates that the highest pitched segments of the syllables show a high degree of regularity: two-thirds of 343 syllables in conversational speech show either the same pitch in successive syllables or regular shifts of 17 vibrations per second or integral multiples of 17 vibrations, 70 per cent being shifts of 17 or of 34 vibrations per second. This suggests that these two figures represent a standard range of pitch (accentual) variation, the latter between low and high pitched syllables, and the former between low and mid or mid and high pitched syllables, within the range of voice of the two subjects.

This paper reports the results of a study whose purpose was to determine whether any regular variations in tonal accent distinguishing the low, mid, and high pitches in the language of two subjects who speak the Tokyo dialect. It is, accordingly, not directly concerned with the traditionally accepted rules of tonal accent, for these are based upon the pronunciation of words in isolation.

Both subjects are natives of Tokyo. The material consists of brief conversational sentences, like the following:

A. *Konnichiwa. Ikaga desu ka?* (Good day. How are you?)

B. *Arigato(o) gozaimasu. Okagesama de genki desu.* (Thank you, I am fine. And you?

A. *Arigato(o) gozaimasu. Aikawarazu genki desu. Kanai wa sukoshi kaze o hikimashita ga mo(o) naorimashita.* (Thank you, I am fine as usual. My wife caught a slight cold, but she is nearly well).

B. *Sa(a), shitsurei itashimasho(o). Sayonara.* (Well, I must be going. Good bye.)

The material comprises 343 syllables. The conclusions were reached by means of data obtained from kymograph records. Throughout this paper we follow general usage in regard to the two terms "pitch" and "frequency", using the word "pitch" to refer to acoustic effects and "frequency" to refer to the number of air vibrations.

In endeavoring to correlate the pitch accent as acoustically per-

ceived with its physical basis, three lines of study were pursued, in the expectation that this basis would be found in one or more of them. Comparison was made:

1. Of the difference between the average pitches of successive syllables;
2. Of the difference between the highest note of each syllable and the highest note of the following syllable;
3. Of the shifts in pitch that take place at the juncture of syllables.

Our first step was to calculate the average pitch of each syllable. The average pitch of the first syllable in each sentence was then compared with the average pitch of the second, that of the second with the average pitch of the third, and so on, thus making it possible to plot a tonal curve for each sentence. We then determined the relative frequency of occurrence of three types of pitch: (1) falling pitch, in which the pitch falls from one syllable to the next; (2) rising pitch, in which the pitch rises from one syllable to the next; (3) level pitch, in which the pitch remains unchanged through two or more successive syllables. The percentages of frequency of occurrence were: falling pitch, 65 per cent; rising pitch, 33 per cent; level pitch, 2 per cent. These percentages show twice as many falling pitches as rising pitches, and this fact corresponds roughly to the acoustic impression. But the comparison of average pitches did not show any regular, standard differences in tonal variation between the low, mid, and high pitches of successive syllables. Furthermore, it showed that only 2 per cent of the syllables have level pitch, whereas the ear recognizes a large number of such syllables. These results justify the conclusion that the characteristics of Japanese sentence intonation are not determined by the average pitches of syllables.

The second step consisted of a comparison of the highest note in each syllable with the highest note in the following syllable, and so on throughout each sentence. The following procedure was followed to determine which note was the highest in each syllable. The voiced portions of each syllable were divided into segments of 0.03 sec. each, since this is the smallest segment the average pitch of which can be calculated with a high degree of accuracy. (In some instances only 0.02 sec. remained at the end of a syllable). By "highest note" above mentioned is meant that segment which shows the highest

average pitch in each syllable. The results here are more significant. It became apparent at once that the proportion of rising, falling, and level pitches corresponded to the acoustic impression of the sentence intonation. Again the total number of falling pitches was nearly twice the number of rising ones; the number of level pitches was 84. The corresponding percentages are: falling pitch, 48 per cent; rising pitch, 27 per cent; and level pitch, 24 per cent. As would be expected, the percentages of falling pitches and rising pitches correspond roughly to those obtained in the first study. Especially significant in this second study is the fact that the proportion of level-pitch transitions between syllables (24 per cent) agrees with the aural impression. This is in sharp contrast with the frequency, only 2 per cent, obtained on the basis of the average syllable pitch.

We may now consider whether any regular, standard differences in frequency are discoverable in the numerical differences in air vibration marking the shifts in pitch between the highest notes of each pair of syllables. A tabulation of such numerical differences reveals that of the 343 syllables considered 239 show either level junctures or shifts of 17 vibrations or of multiples of 17 vibrations, that is, 34, 51, and so on. Furthermore 17 and its first multiple, 34, constitute nearly 70 per cent of the 239 instances. The speech of one of the subjects, whose pronunciation is more regular, shows a still more striking regularity, namely, 80 per cent of levels and of shifts of 17 or its multiples; this leaves a scattering of junctures that amount to only one fifth of the entire number. When we examine these last junctures, we find that the great majority of them contain the vowels *u* and *i*, which are often omitted entirely in pronunciation and which commonly tend to be faintly pronounced and therefore to be pronounced at lower pitches. Other exceptions are explained by the fact that the semivowels *n*, *m*, and *r*, all of which in Japanese approximate vowels in pronunciation, merge in some instances with a preceding or a following vowel in both pronunciation and acoustic effect. These cases are difficult to interpret, because it is frequently impossible, owing to overlapping muscular movements of the speech organs involved, to distinguish in kymograph or any other instrumental records where the semivowel ends or the vowel begins, and whether two syllables are involved or three.

If we turn back now to the recurring shifts of 17 vibrations and of its multiples, 34, and so on, the thought immediately suggests itself that 17 vibrations may represent the difference between the high and the mid or between the low and the mid tones, and that 34 vibrations, being twice as great, may represent the difference in frequency between low and high. Traditionally the shift from high to mid as well as the shift from mid to low has been stated to average slightly more than the interval between two musical notes, and the difference between low and high twice that amount. The conversational voice of one of our subjects ranged through 150 (one instance of 133) to 235 vibrations (mid baritone range), the other through 183 (one instance of 150) to 315 vibrations (upper baritone range). Within this range of the musical scale the difference in frequency between two successive notes varies from 16 to 32 vibrations, the average (if one is justified in using the average difference as a measure of the difference of Japanese pitch) is 22 vibrations. Perhaps this may be considered slightly more than one musical note. Close agreement between our findings and the traditional rule is not to be expected, since we are dealing with the accent of connected discourse, whereas the traditional view is based upon the pronunciation of individual words. It would be interesting to learn whether a bass voice would show standard shifts in frequency materially less than 17 and 34.

Why do these characteristic shifts not appear in the first phase of our study? Evidently the characteristic high notes which distinguish the intervals of pitch accent, are submerged in the average frequency of the syllable, and therefore no standard difference can be deduced from the comparison of average frequency.

The third step in this study dealt with the juncture of syllables. During the utterance of any two successive sounds, the movements of the speech organs productive of any one sound ordinarily overlap those of the next succeeding sound. Thus it is not possible to mark with precision the limits of syllables. Accordingly in this study the phrase "juncture of the two syllables" is taken as a transitional period of 0.06 sec. including the last 300th of a second of one syllable and the beginning 300th of a second of the next following syllable. Since the average length of the syllable in our material is about 0.15

sec., this period of transition is about $^1/_5$ the total length of the two syllables.

The data of this last comparison confirm the findings of the second in two particulars: (1) 17 vibrations and its multiples (34, etc.) reappear as standard shifts between the low, mid, and high tones in 75 per cent of all our cases; (2) the proportion of falling pitch to rising pitch is practically identical in the second and third phases of our study, being 64 per cent to 36 per cent in one and 70 per cent to 30 per cent in the other. This means that the falling pitches are about twice as many as the rising ones. This is explained by the fact that in Japanese the fall from high to low goes through one or more syllables of middle pitch instead of falling abruptly from high pitch to low pitch.

In view of all the above facts, we feel justified in concluding that frequencies of approximately 17 and 34 vibrations per second represent quite closely the acoustically prominent tone shifts from syllable to syllable in the Tokyo dialect of our subjects. Although this study is based upon the pronunciation of only two subjects, the writer's experience leads her to believe that their sentence intonation represents fairly closely the pronunciation of the women who speak the Tokyo dialect.

University of Michigan,
Ann Arbor, Michigan

HIDE HELEN SHOHARA