Ann.Bull.RILP(1989)No.23, 141-152

PRODUCTION OF THE FOUR CHINESE TONES BY AN ESOPHAGEAL SPEAKER

Hajime Hirose

Introduction

Our earlier study on the communicative ability of esophageal speakers in terms of the production of linguistic contrasts exclusively dealt with the voicing distinction for consonants in Japanese as one of segmental aspects of esophageal speech. Through perceptual, acoustic and fiberoptic investigations, it was revealed that skilled esophageal speakers were able to regulate their neoglottis to achieve relatively accurate control for the voicing distinction in Japanese1).

It has been widely accepted that another linguistic aspect of speech, i.e. the prosodic aspect of speech, plays an important role in human communication among normal as well as alaryngeal subjects. As one piece of the evidence that esophageal speakers can achieve prosodic control, Gandour, Weinberg and Petty2) reported that esophageal speakers were able to produce systematic changes in fundamental frequency (F0) and intensity to a considerable extent in relation to lexical stress in American English.

In a study of the ability of esphageal speakers in producing Japanese accent contrasts, Takahashi and his colleagues3) reported that the accent pattern was correctly identified by normal listeners in 60.8 % of their perceptual tests, and that an F0 change was the primary cue for the identification of the accent pattern. They also noted that the rising pitch accent type was more correctly perceived than the falling type.

In tone languages like Chinese, it is known that the tonal patterns are defined at the word or syllable level and regulated in the form of F0 changes in the acoustic domain. According to Eady4), the speech of normal Chinese subjects is characterized by larger rates of F0 change compared to American English speakers. These facts raise the question of whether Chinese esophageal speakers would show more difficulty in producing tonal contrasts than speakers of other types of languages.

As for the ability of esophageal speakers to produce tonal contrasts, Gandour and his group5) examined two esophageal speakers of Thai, which has five contrastive lexical tones. The performance of their subjects in terms of their production of the tonal contrasts did not reach the level of proficiency of normal speakers.

Recently, we had an opportunity to examine a Chinese female alaryngeal speaker of the Beijing dialect who was visiting Japan for a short period for a continuing training program in esophageal speech sponsored by the Asian Association of the

Laryngectomized.

In standard, or Mandarin Chinese, there are four lexical tones. These lexical tones are referred to as tone-1, tone-2, tone-3 and tone-4, respectively. The basic pitch patterns of the four tones obtained by spectrographic analysis were described by Chao(6), and Shen et al.(7) as shown in Fig. 1-a. Later, Chuang and his group(8) made an extensive acoustical study of Chinese monosyllables and described the basic patterns shown in Fig. 1-b. According to Chuang et al., the essential features of the pitch patterns of the four tones are as follows.

tone-1: flat pattern, mid-high

tone-2: rising patterns from mid-low to mid-high

tone-3: long pattern with a low flat beginning and rising end

tone-4: short falling pattern from high to low

The aim of the present study was to investigate the perceptual and acoustic characteristics of the four Chinese tones produced by the Chinese esophageal speaker and to compare her performance to that of the normal Chinese speakers reported in literature. A perceptual study was first performed to examine the pattern of confusion that native listeners of Chinese would make in comprehending the tonal pattern of the esophageal speaker. A computer analysis of recorded speech samples was then made to invesitigate their acoustic nature with special reference to F0 contours.

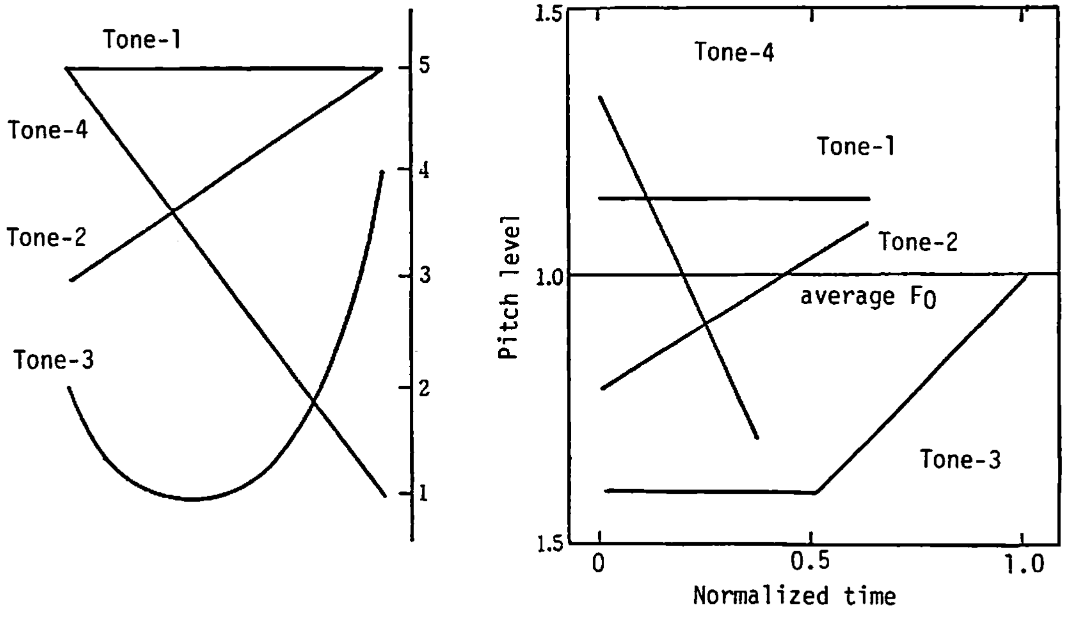
Procudures

1. Experimental subject

The subject was a 40-year-old Chinese female of the Beijing dialect, born in Beijing in 1948. She underwent total laryngectomy for carcinoma of the larynx in 1986 at the age of 38. She has been practicing esophageal speech since and working as an employee of a local meteorological office near Beijing. She visited Japan for 3 months in early 1988 to continue her training in esophageal speech at Ginreikai, the largest association of the laryngectomized in Japan.

2. The perceptual test

The subject was required to read the 16 meaningful Chinese words shown in Table 1, each of which was indicated to the subject in written form on a card in random order, with pauses between words. All the words appeared three times each so that total of 48 utterances are produced. No frame sentence was used.



a)

b)

Fig. 1 Typical pitch patterns of the four different Chinese tones proposed by Chao (a) and Chuang et al. (b)

	1	2	3	4
/ma/	媽	麻	馬	罵
/mao/	猫	毛	卯	帽
/fa/	髪	乏	法	珐
/fen/	分	坟	粉	奋

Table 1 The 16 test words used in the present experiment (Arabic number indicates tone type).

Recordings were made in an anechoic room using a unidirectional microphone, and recorded samples were used in the subsequent identification test.

Six native Chinese students of the Beijing dialect participated in the identification test performed in an unechoic room. Each listener was provided with set of answer sheets containing the closed set of stimulus words for each trial written in Chinese characters. Listeners were requested to identify each stimulus by marking the word on their answer sheet corresponding to their judgment. All the listeners were unfamiliar with esophageal speech, and the purpose of the test was not explained beforehand.

	I	II	III	IV
I	44	2	2	0
II	2	29	16	1
III	1	1	46	0
IV	14	0	6	28

ERROR: 15.6%

Fig. 2 Confusion matrix for the test samples (6 judges).

3. Acoustic analysis

The 48 recorded samples used in the perceptual test were analyzed to obtain their F0 contours using a specially designed computer program. The maximum and minimum F0 values and the duration of each sample were also obtained.

Results

1. Results of the perceptual experiment

Figure 2 shows the results of the perceptual experiment in the form of a confusion matrix in which 288 data (16 test words x 3 repetitions x 6 judges) are included. As a whole, there is a relatively small percentage of confuion in the identification of tones (45/288 = 15.6%). A high rate of confusion can be noted in the frequency with which tone-2 was judged as tone-3 (16/72 = 22.2%) and tone-4 was judged as tone-1 (14/72 = 19.4%). There was no significant difference in the rate of confusion with reference to the type of test word.

2. Results of the acoustic analysis

As for the overall ability of the present subject at F0 regulation, it was revealed that her average speaking pitch during the recording was 169 Hz and this value is higher than that of the majority of esophageal speakers. Her highest F0 value was 300 Hz and the lowest was 96 Hz. The average pitch elevation in tone-2 and tone-3 was 73 Hz, while the average pitch drop in tone-4 was 76 Hz.

Figures 3-6 show the F0 contours for all the recorded samples grouped by tone types.

In each figure, there are 3 samples for each of the different test words. When a confusion occurred for a given sample, it is marked in the figure by an asterisk. The fraction attached indicates the rate of confusion, in which the numerator indicates the number of Judges out of six (the denominator) showing the confusion. When the numerator is larger than 3, i.e. if more than a half of the judges showed the confusion, the sample is marked by two asterisks. The type of confusion is also indicated by a Roman number corresponding to the tone type. For example, in Figure 3, the first sample for the test word /fa/ with tone-1 was confused by 4 judges, 2 of whom judged it as type-2. The other 2 judges said it was type-3.

The results indicate that confusion occurred in a relatively limited number of samples. The nature of this confusion will be discussed in the following section.

In order to investigate the relationship between the type of confusion and the duration of each sample, a relative duration of each sample was obtained. Namely, for each group of test words

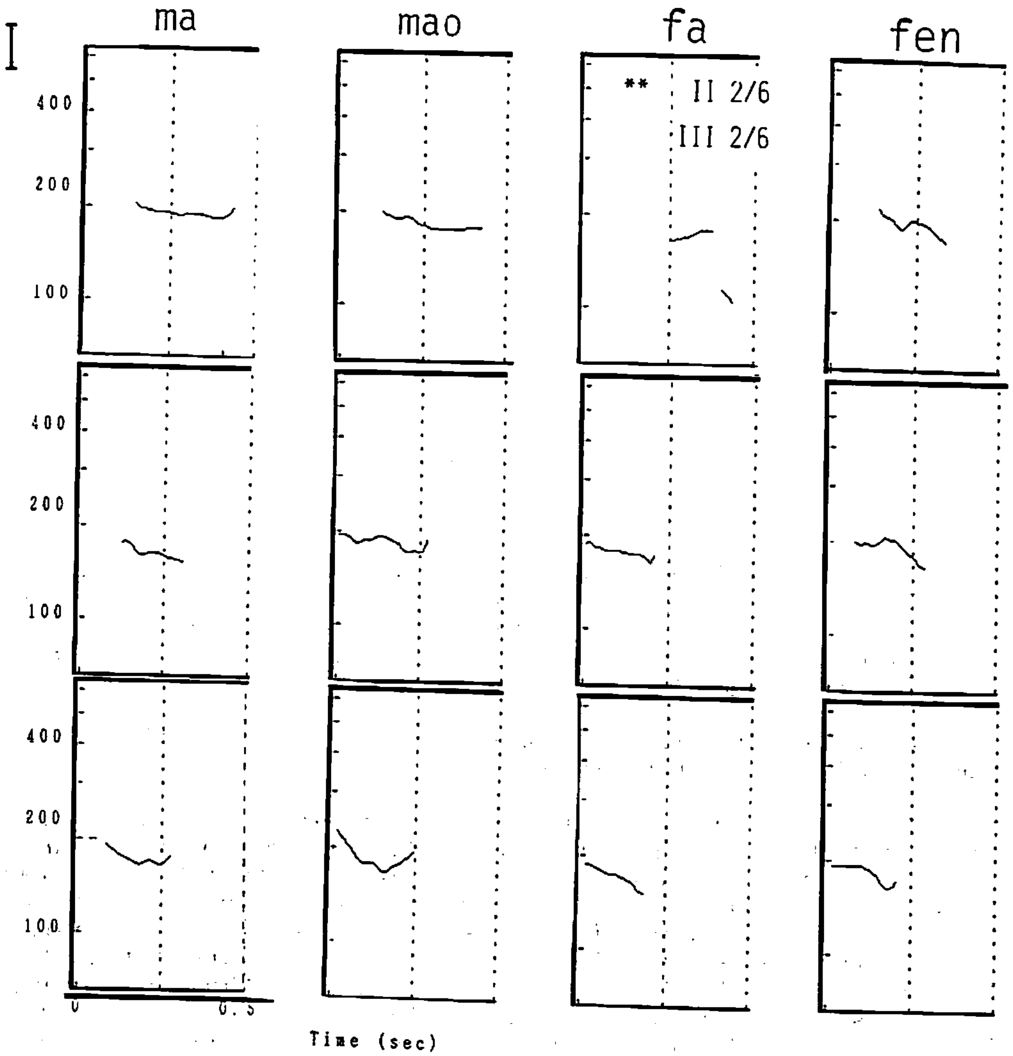


Fig. 3 F0 contours of the samples uttered to produce tone-1.

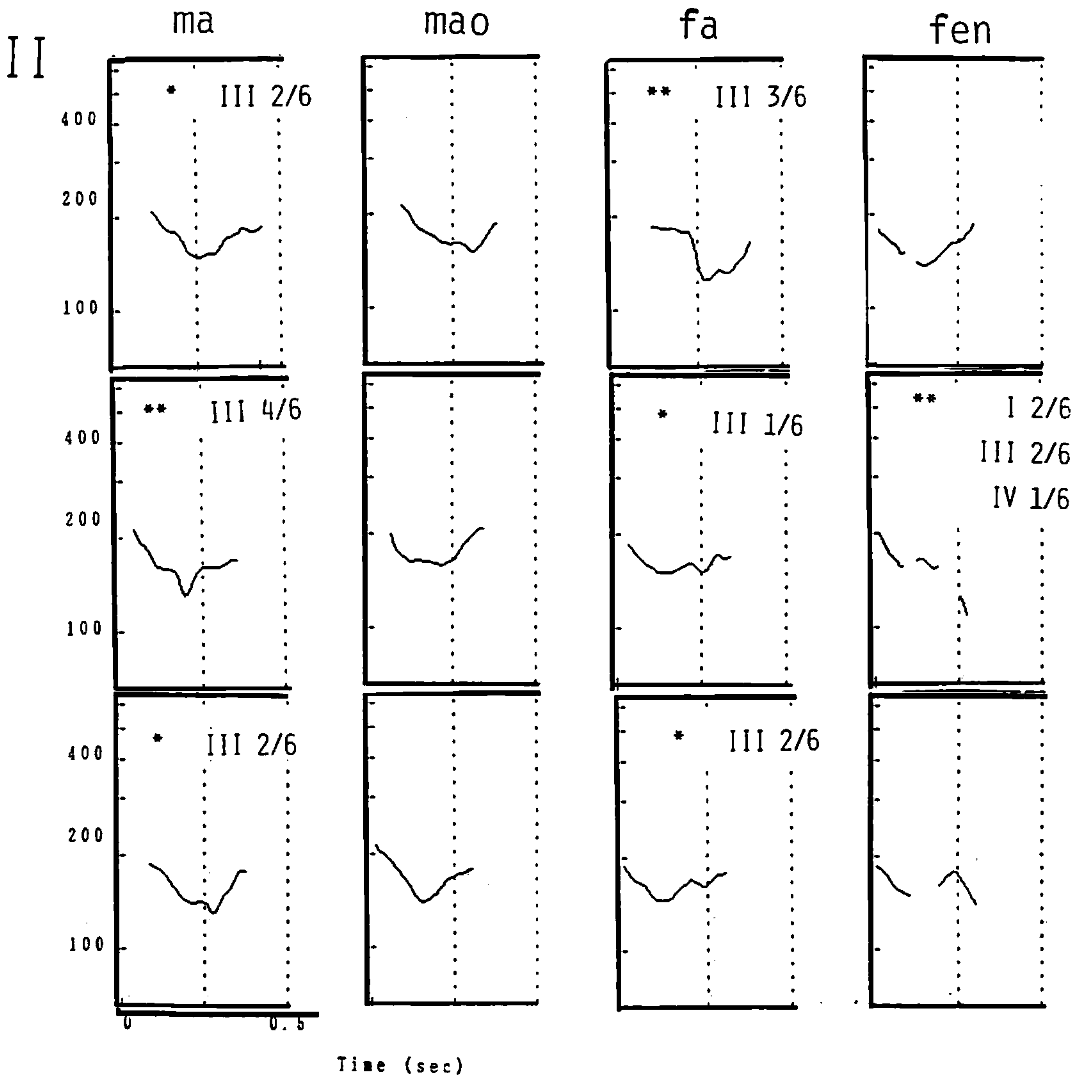


Fig. 4 F0 contours of the samples uttered to produce tone-2.

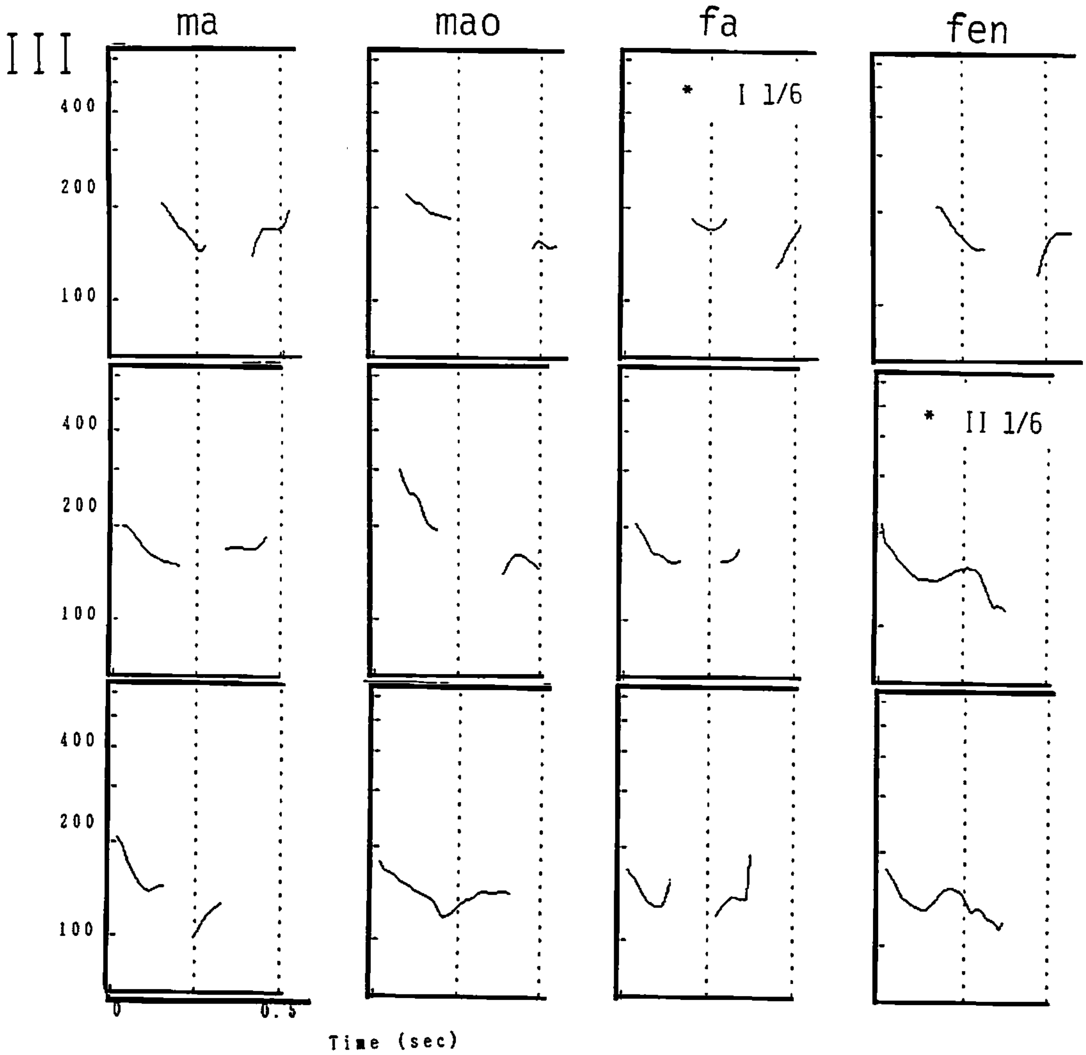


Fig. 5 F0 contours of the samples uttered to produce tone-3.

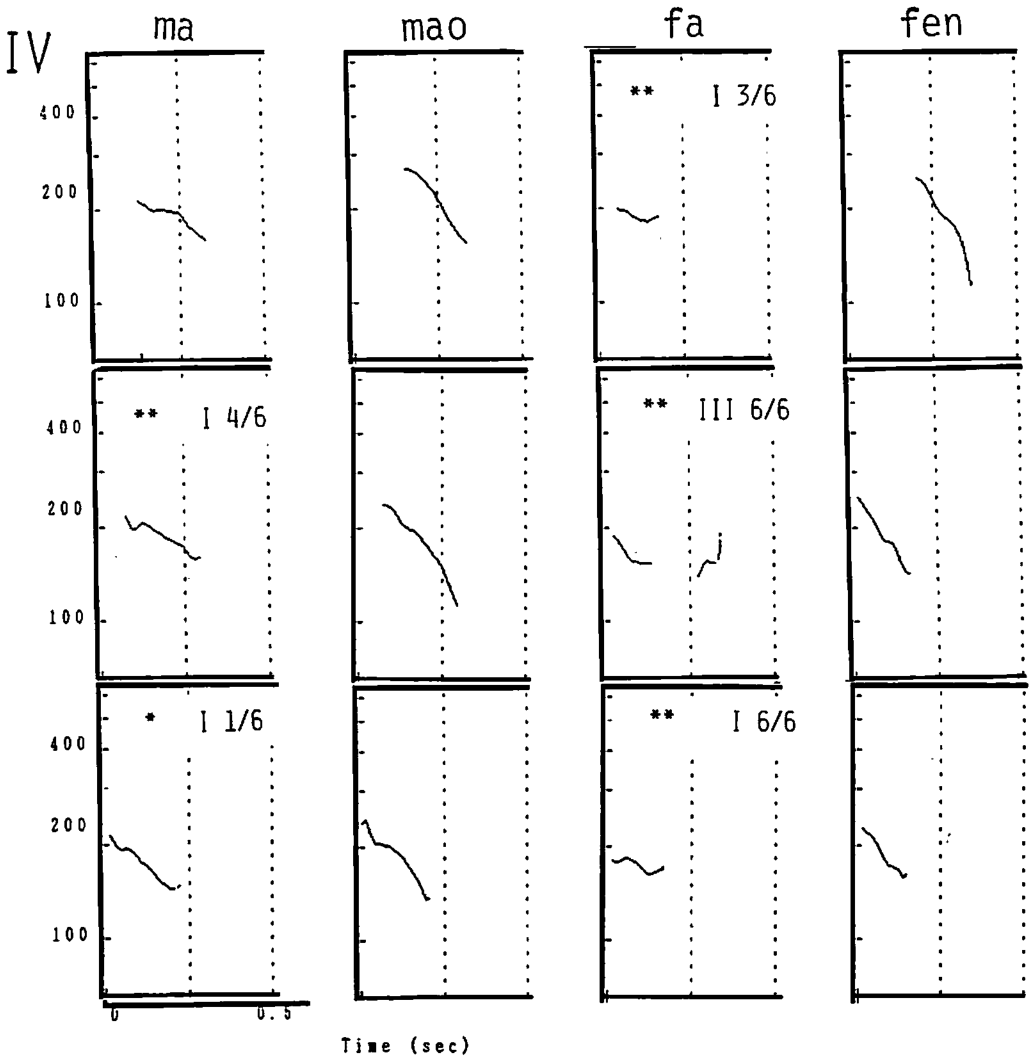


Fig. 6 F0 contours of the samples uttered to produce tone-4.

having the same phonemic structure, the duration of each sample was normalized by the average of all samples belonging to the same group. Then, all the samples were grouped by each tone-type and plotted in Figure 7, in which confused samples are indicated by filled circles. It was shown that, except for one specific sample for tone-4, there was no remarkable difference in the relative duration between the confused and the correctly identified samples. This would indicate that the duration of the test words was not a primary cause of confusion, and that the durational characteristics for each tone type were well maintained in the speech production in this particular subject.

Comments

As described by Chuang et al.[8], each of the four Chinese tones is acoustically characterized by its F0 contours, although durational characteristics may also be taken into consideration.

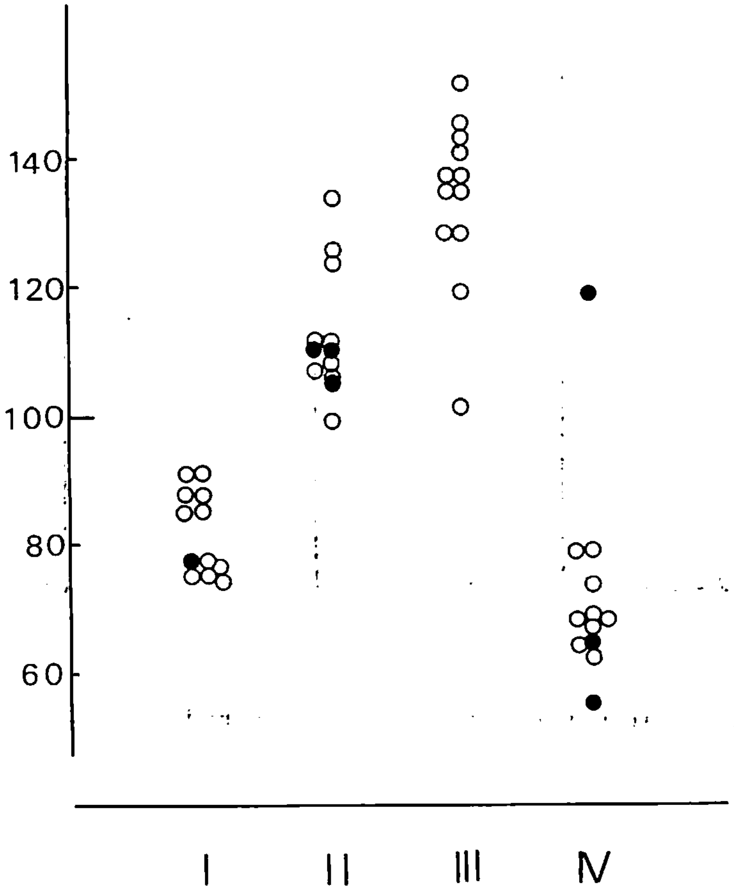


Fig. 7 Distribution of relative duration of all the samples grouped by the intended tones.

When we closely examined each of the samples perceptually confused in the present experiment with reference to its acoustic nature, most of the confusions appear to be caused by an atypical F0 contour of the samples.

For example, when the F0 contour stayed more or less flat despite the subject's intention to produce tone-4, the sample was perceived as tone-1. Although the subject intended to produce tone-4, the sample was perceived as tone-3 when there was a short final rise in the F0 contour preceded by a sharp descent.

A relatively high confusion rate between tone-2 and tone-3 has been noted even in normal Chinese subjects. Chuang et al.8) reported that the F0 patterns of tone-2 and tone-3 obtained from normal speakers of Chinese resembled each other and that the resemblance resulted in a confusion rate of approximately 10 %. According their perceptual experiments, the error rate for identifying tone-3 as tone-2 was almost twice as high as that for the opposite confusion. They suggested that neither the falling nor the concave pattern alone can be a perceptual cue in discriminating tone-3 from tone-2, but that a continuous low-flat period is an important cue for tone-3.

In most of the present samples of tone-3, there is often a period with a low F0 level associated with a marked decrease in intensity after the initial pitch fall so that the F0 contour appears to be interrupted. The period is then followed by a final pitch rise. This pattern seems to be produced intentionally for characterizing tone-3 by the present subject to realize the continuous low-flat period before the final elevation. This would explain why the confusion of tone-3 with tone-2 was rarely noted in the present experiment.

On the other hand, the nature of the confusion of tone-2 as tone-3 is not always clear. Chuang et al.8) claimed that the acoustic cues characterizing tone-2 seemed quite complex, and that the rate of confusion of tone-2 as tone-3 appeared to increase when the initial pitch descent became more marked. A similar tendency was noted in the present experiment in that the confused samples of tone-2 often showed a relatively marked pitch descent immediately prior to the final rise, although this is not yet conclusive.

The present study clearly indicates that at least this particular esophageal speaker is able to regulate the F0 contour of her speech almost satisfactorily in an experimental situation and that perceptual confusions mostly occur when samples have an atypical F0 contour.

In general, however, esophageal speakers show a difficulty in F0 regulation. Snidecor9) reported that the average speaking pitch of his esophageal speakers was 63.3 Hz in both males and females, and that their pitch variation was hardly perceivable. According to Sawada10), the speaking pitch range in his 7 esophageal speakers was between 55 and 130 Hz, and none of these could produce a pitch variation beyond 1 octave. The ability of the present subject in pitch regulation appears to be superior to these previous reports.

The mechanism for pitch regulation in esophageal speech is not obvious. In our preliminary experiment on Japanese accent production, an increase in intraesophageal pressure was noted for the production of accented mora, suggesting there is an effortful control for "markedness". Esophageal speakers often claim introspectively that they tense their neck region and attempt to increase abdominal pressure for the production of accented morae. Although further physiological study is necessary, the present study would suggest that the training of esophageal speech may lead to a considerable achievement of pitch regulation.

References

1) Hirose, H., M. Sawashima and H. Yoshioka: Voicing distinction in esophageal speech -perceptual, fiberoptic and acoustic studies. Ann. Bull. RILP., 17, 187-199, 1983.
2) Gandour, J., B. Weinberg and S.H. Petty: Production of lexical stress in alaryngeal speech. Folia phoniat., 37, 279-286, 1985.
3) Takahashi, H., M. Yoshida, S. Tsumura, Y. Sakamoto and T. Yamasaki: Study on the changeability of fundamental frequency of esophageal speech and artificial larynx. Jpn. J. Logoped. Phoniat., 20, 147-152, 1979.
4) Eady, S.: Differences in the F0 patterns of speech: Tone language versus stress language. Language and Speech, 25, 29-42, 1982.
5) Gandour, J., B. Weinberg and S.H. Petty: Tone in Thai alaryngeal speech. J. Speech Hear. Dis., 53, 23-29, 1988.
6) Chao, Y.R.: Concise Dictionary of Spoken Chinese. Cambridge, 1947.
7) Shen, Y., J.C.Y. Chao and C. Peterson: Some spectrographic light on Mandarin tone-2 and tone-3. The Study of Sound, 9, 265-314, 1961.
8) Chuang, C.K., S. Hiki, T. Sone and T. Nimura: Acousitical features of the four tones in monosyllabic utterances of standard Chinese. J. Acoust. Soc. Jpn., 31, 369-380, 1975.
9) Sawada, K.: An experimental study of alaryngeal speech. Otol. Clinic. (Kyoto), 52, 42-72, 1959.
10) Snidecor, J.C.: On esophageal speech. Jpn. J. Logoped. Phoniat. 9, 5-8, 1968.