J. Acoust. Soc. Jpn. (E) 9, 6 (1988)

On the listening tests using natural speech of 8 Korean vowels

PACS number: 43.71.Es

H.-Y. Chung, S. Makino, and K. Kido
*Research Center for Applied Information Sciences,
 Tohoku University,
 2-1-1, Katahira, Sendai, 980 Japan
 (Received 14 April 1988)*

1. Introduction

In the analysis of isolated Korean vowels using formant frequencies, the vowel distributions on F_1 - F_2 plane show heavy overlappings 1) between /e/ and /ɛ/ for all 20 speakers, and 2) between /ə/ and /ʌ/ for Kyungsangdo speakers.¹⁾

It is very interesting problem to check whether Korean listeners also make such confusions in the perceptions as observed in the utterances. And if the relationships between utterances and perceptions are made clear, we can easily find out the factors causing confusions in utterances or in perceptions. Furthermore, if the recognition score by human for the vowels is given, it would be very useful measure for evaluating a recognition method.

In order to investigate these points, we conducted listening tests for the above-mentioned utterances of isolated vowels. Based on the listening tests, we define the perceptual domain; the regions of correct perception of the listeners for a vowel, for each of the 8 vowels, and make clear the relationships between the utterance distributions and the perceptual domains.

2. Listening Test

The two different dialects, Seoul and Kyungsangdo, were also chosen for listening tests because the Seoul dialect is the standard Korean and the Kyungsangdo dialect is the secondarily used dialect.

The 8 vowels were uttered two times by each seven speakers from the two dialects so that our speech data consisted of a total numbers of 224 vowels ($8 \times 2 \times 7 \times 2 = 224$). In the tests the vowels of the speech data were presented three times to five subjects from Seoul and five from Kyungsangdo in random order. Each listener heard one vowel every second and wrote down the perceived vowel's name.

3. Perceptual Domain

For the investigation of the perception of vowels

and for the graphic representation of the perceived vowels we would like to define a perceptual domain in such way, that a sample vowel which was correctly perceived by all listeners has a strong influence in the perceptual domain while those vowels which were rarely perceived correctly has a weak influence. So, we defined a weight W_{ij} for each sample vowel ($j=1, 2, \dots, 224$) with the ratio of the correctly perceived vowels as the i -th vowel ($i=1, 2, \dots, 8$) to the numbers of total vowels ($15=5$ listeners $\times 3$ times) in the sample. In the same way, we define a weighted mean vector u_{ik} of the k -th formant frequency of the samples perceived as the i -th vowel

$$u_{ik} = \frac{1}{\sum_{j=1}^N W_{ij}} \sum_{j=1}^N W_{ij} X_{jk}$$

where X_{jk} is the k -th formant frequency of the j -th sample, and $N=224$. The weighted covariance vector C_{ikl} of the i -th vowel is given by

$$C_{ikl} = \frac{1}{\sum_{j=1}^N W_{ij}} \sum_{j=1}^N W_{ij} (X_{jk} - u_{ik})(X_{jl} - u_{il})$$

Based on these weighted mean vectors and weighted covariances, we plot 1.64 sigma ellipsoids which include 90% of the samples of a population. We call each of the ellipsoid the 'perceptual domain' of the vowel.

4. Listening Results and Perceptual Domains

Confusion matrices of the listening test for six vowels are shown in Table 1. Because the two vowels /a/ and /i/ showed no confusions with the other vowels and vice versa, we have excluded the scores for the two vowels from the Table 1.

If we regarded /ɛ/ and /e/ as one vowel the listening tests showed the following results: presenting Seoul speakers' utterances the vowels were perceived correctly by 96.8% of the Seoul listeners, 90.1% of the Kyungsangdo listeners; presenting Kyungsangdo speakers' utterances the vowels were perceived correctly by 8.13% of the Seoul listeners, 83.2% of the Kyungsangdo listeners.

By comparing utterance distributions with their perceptual distributions, we can understand why the two dialectal listeners make confusion one vowel with the other vowels in the listening tests. The utterance distributions and the perceptual domains of the vowels of the two dialects are shown in Fig. 1 (a)~(d). The four confusion matrices shown in Table 1 correspond to the plots in Fig. 1 (a)~(d) respectively.

As we can see in Fig. 1, the vowels /ɛ/ and /e/ of both dialects show heavy overlappings in utterances, in

Table 1 Results of listening test for 6 vowels (%).

Uttered	Listened											
	by 5 Seoul listeners						by 5 Kyungsangdo listeners					
	u	o	ɛ	ə	e	ʊ	u	o	ɛ	ə	e	ʊ
by 7 Seoul speakers												
u	94.9	5.1					96.3	3.0				0.7
o	0.7	98.6		0.7			5.9	94.1				
ɛ			41.5		58.5				53.3		46.7	
ə				92.6		4.4		9.6		74.9		9.6
e			26.6		73.4				48.2		51.8	
ʊ	3.0			5.2		91.8				30.4		69.6
by 7 Kyungsangdo speakers												
u	56.7	31.3		4.0		8.0	67.4	18.7		6.6		7.3
o		68.7		31.3			0.7	84.0		10.0		1.3
ɛ			43.3		56.7				68.0		32.0	
ə	0.7	0.7		90.0		8.6		2.7		71.3		26.0
e			49.3		50.7				64.0		36.0	
ʊ				67.3		32.7				61.3		38.7

Fig. 1 Perceptual domains and utterance distributions in Seoul and Kyungsangdo people.

perceptions and between utterance distributions and perceptual domains for both dialects.

Figure 1(a) shows Seoul's perceptions for Seoul's utterances. All of the vowels are separated from each other in utterances, except /ɛ/ and /e/, but slight overlappings are seen between /u/ and /o/, and between /o/ and /ə/ in perceptions.

In Fig. 1 (b) which shows Kyungsangdo's perceptions for Seoul's utterances, slight overlappings are seen between utterance distributions and perceptual domains for the vowels /o/ and /u/, /ə/ and /o/, and between perceptual domains of /u/, /o/, /ə/ and /w/. Perceptual domain /w/ of Kyungsangdo listeners overlaps heavily with their /ə/ and overlaps slightly with /ə/ of Seoul's utterances. So, perfect perceptions for these vowels are not expected.

Figure 1(c) shows Seoul's perception for the Kyungsangdo's utterances. A greater part of /w/ of Kyungsangdo's utterances overlaps with the perceptual domain of /ə/ of Seoul. It is thus difficult for Seoul listeners to distinguish between the /w/ and /ə/ vowels of the Kyungsangdo speakers as can be seen in Table 1. When we investigate Kyungsangdo's utterance distributions of /u/ and /o/, /u/ is located between the perceptual domains /u/ and /o/ of Seoul listeners, and /o/ is located between /o/ and /ə/ of Seoul's perceptions, causing Seoul listeners to confuse /u/ with /o/ and /o/ with /ə/. Utterance distributions and perceptual domains of Kyungsangdo dialect are shown in Fig. 1(d). The perceptual domains of /ə/ and /w/ overlap each other and include their utterance distributions. By comparing the Fig. 1(d) with the Fig. 1(c) we can see that perceptual domains of Kyungsangdo listeners have a higher F_1 for /o/ and a higher F_2 for /ə/ than that of

Seoul listeners do, leading to less overlappings with utterances for /u/ and /o/ of their dialect, and leading to reduced confusions for their dialect in the listening test.

5. Summary

From the above investigation, perceptual domain for each vowel are well agreed with the results of our listening test. In both utterances and perceptions, /ɛ/ and /e/ are not considered to be separated in both Seoul and Kyungsangdo dialects. For Kyungsangdo dialect, heavy overlapping between /ə/ and /w/ is seen in both utterances and perception, so it can be concluded that Kyungsangdo people pronounce and perceive the two vowel without distinction. Utterance distribution of Kyungsangdo people has a slightly higher F_1 for /o/ and their perceptual domain has a little higher F_1 for /o/ and a little higher F_2 for /ə/ compared to the Seoul's perceptual domains. Therefore, they perceive of their own dialect with lower confusion than they do for Seoul's utterances.

In human perception of speech the ability of adaptation to speakers or dialectal characteristics is well-known. It is very interesting to investigate, by presenting another combination of speech samples to human listeners, in what extent perceptions is improved by the adaptation effect. This comparative investigation will be the subject of our further studies.

Reference

- 1) H. -Y. Chung, S. Makino, and K. Kido, "Analysis and recognition of Korean isolated vowels using formant frequency," *J. Acoust. Soc. Jpn. (E)* **9**, 225-232 (1988).