Korean talkers' cue weighting perception strategies in perceiving English /ɑ/, /ɔ/ and /ʌ/ in comparison with American talkers'*

Soonhyun Hong
(Inha University)

Hong, Soonhyun. 2013. Korean talkers' cue weighting strategies in perceiving English /ɑ/, /ɔ/ and /ʌ/ in comparison with American talkers'. *Studies in Phonetics, Phonology and Morphology* 19.3. 529-554. Korean university learners of American English have difficulty differentiating between English /ɑ/, /ɔ/ and /ʌ/ as well as between /i/ and /ɪ/, /ɛ/ and /æ/ and /ʊ/ and /u/. Yang (2010) and Hong (2011, 2012) discussed extensively Korean talkers' perception of the latter three vowel pairs only. This paper, on the other hand, focuses on Korean talkers' perception of English /ɑ/, /ɔ/ and /ʌ/, among which Korean talkers have difficulty differentiating. This paper, using a forced-choice identification test with spoken hVd stimuli in Hillenbrand et al. (1995), hypothesizes that Korean and American talkers use different cues with different weighting despite the existence of corresponding similar Korean vowels (Morrison 2002, Escudero and Boersma 2004, Holt and Lotto 2006, Ylinen et al. 2009). Based on the measured duration, F0, F1, F2 and F3, stepwise Discriminant Analysis, a multivariate analysis, was waged on the identification performance for English /ɑ/, /ɔ/ and /ʌ/ by American and Korean talkers, deriving two discriminant functions. These functions are latent variables (i.e., linear functions of weighted cues) and were used to plot vowel stimuli and vowel group centroids in perceptual vowel space with two latent variables. As a result, a given vowel stimulus will be plotted closest to its vowel group centroid. By comparing Korean and American talkers' models, it was found that Korean talkers use different strategies in perceiving the three vowels in question from American talkers as to how those latent functions were composed of for English vowel perception. The two latent variables explained 98.3% of American talkers' identification performance while 71.3% of Korean talkers' identification. These results suggest that Korean talkers use different perception strategies in differentiating between the three vowels from American talkers and their perception strategies are not stable. **(Inha University)**

Keywords: vowel identification, perception of /ɑ/, /ɔ/ and /ʌ/, Discriminant Analysis, cue weighting, latent dimension, perceptual vowel space

1. Introduction

Perceiving L2 speech sounds poses difficulties when they do not have corresponding sounds in L1 and as a result, dissimilarities arise in the use of phonetic cues in L1 and L2 phonological inventories.

Ylinen et al. (2009) reviewed L2 acquisition literature and pointed out three possible cases as to L1 talkers' perceptual difficulties in terms of

* This work was supported by Inha University Research Fund. Many thanks go to three anonymous reviewers for their insightful comments and suggestions, though all errors are mine.

perceptual roles of cues. First, if two vowels in L2 have only one corresponding vowel in L1, then L1 and L2 may use the same cue but the critical values of the cue, for example, in the acoustic vowel space of F1 and F2, may not be the same (Flege 1988). Second, L2 vowels may be perceived using cues which were never used in perceiving a corresponding similar L1 sound. American talkers rely more on formant cues than duration in differentiating between, for example, English /i/ and /ɪ/ (Hillenbrand et al. 2000). On the other hand, Korean talkers crucially use duration (Ingram and Park 1997). Third, multiple cues may have to be used for the perceptual distinction of L2 vowels. Namely, several acoustic cues may contribute simultaneously but unequally (Flege and Hillenbrand 1986, Holt and Lotto 2006). As a result, these cues will have to be considered unequally in category perception.

Furthermore, L2 talkers may ignore critical cues but failingly struggle to pick up unreliable cues instead. For example, Japanese talkers used unreliable F2 values rather than more critical F3 values for /l/ and /r/ distinction (Iverson et al. 2003). Finnish talkers picked up unreliable duration for the contrast between English /i/ and /ɪ/ whereas English speakers use F1 (Ylinen et al. 2009).

Specifically for the perception of L2 vowel categories by L1 talkers, one single cue may not be enough to define category membership. Multiple cues may be involved and different cues may contribute differently to vowel category membership identification. Some cues may contribute more but the other cues less. And some other cues may turn out to be unreliable cues and will never be used. In the literature on cue weighting for vowel perception, most of the focus was placed on binary perception of /i/ and /ɪ/ by L2 talkers in comparison with by English talkers. And one linear function with differently weighted cues was proposed to explain the distinction (Morrison 2002 for Japanese and Mexican-Spanish talkers, Escudero and Boersma 2004, Boersma and Escudero 2005 for Spanish talkers).

In Morrison (2002), the stimuli were manipulated in terms of five steps of vowel spectra (varying but equally spaced combinations of F1, F2 and F3), seven different steps of duration, five steps of plosive closure duration and two different speaking rates and a forced-choice identification was conducted. Discriminant Analysis turned out one linear discriminant function with differently weighted cues for Japanese talkers' perception. Japanese talkers relied more heavily on vowel duration (0.996) than vowel spectra (combination of F1, F2 and F3). In contrast, Canadian English talkers' discriminant function had the most highly weighted vowel spectrum (1.395) but low weighted vowel duration. On the other hand, Escudero and Boersma (2004) focused on the weighting of F1 and vowel duration incorporated in a single function for the perception of manipulated English /i/ and /ɪ/ on F1 and duration simultaneously.

The experimental design of this type begins with stimuli manipulation at equal steps on the combinations of cues. Then L1 and L2 subjects conduct an identification test on these stimuli. Finally, a linear function, i.e., linear combinations of weighted cues, is derived for L1 and L2 talker groups. This function is actually interpreted as another new latent variable or more specifically a latent perceptual dimension along which two groups of vowels (e.g., /i/ and /ɪ/) are represented and compared, just like stimuli of /i/ and /ɪ/ are plotted along the F1 variable in the traditional acoustic vowel chart. This way, it will become possible to compare L1 and L2 talkers' identification strategies for the distinction of the two vowels and to verify whether L2 talkers use reliable cues and proper cue weighting strategies like L1 talkers do. However, if more than three L2 categories are involved, the combinations of all steps of cues for the stimuli will uncontrollably increase and all the finely manipulated stimuli may not be incorporated exhaustively in the identification test.

The Korean talkers' (hereafter, KTs) perceptual difficulties in Korean vowel pairs /i/ vs. /ɪ/, /u/ vs. /ʊ/ and /ɛ/ vs. /æ/ were exhaustively studied in Yang (2010) and Hong (2011, 2012). The difficulties in these vowel pairs lie in the fact that tense/lax distinction (mostly realized along F1) is not properly picked up by KTs. On the other hand, English /ɑ/, /ɔ/ and /ʌ/ are also very difficult to differentiate for KTs. However, KTs' perceptual difficulties for these three vowels do not pattern together with those for the vowel pairs of tense/lax distinction. More cues may be required to differentiate among the three vowels.

We assume in this paper that the cues involved for the distinction of the three English vowels, are steady-state F0, F1, F2, F3 and vowel duration. The analysis in this paper is different from those cue-weighting studies in several aspects. First, the spoken stimuli in the perception experiment are real speech data in the hVd format from the database (Hillenbrand et al. 1995). We selected the vowel stimuli from the database in Hillenbrand et al. (1995). The motivation of the use of natural speech signals of the vowels and the detailed selection procedure will be spelled out in section 2.2. After the vowel identification test was conducted, we analyzed the results, using stepwise Discriminant Analysis with cue measures of the stimuli as the independent variables and subjects' vowel identification clicks, whether correct or incorrect, on the stimuli as the grouping variable. The results showed how vowel identification clicks varied depending on the different combinations of measures of cues. We extracted two latent variables (or latent linear functions of the weighted cues) from the cue variables, which in turn were used as latent perceptual dimensions to plot vowel group centroids and group member stimuli on the two-dimensional perceptual vowel space. The resulting plotted perceptual vowel space looked similar to the acoustic vowel space of F1 and F2. The resulting vowel group plots are clearly separated from one another according to the vowel group. Based on this, it will be argued that 1) perceptual vowel space with two latent variables or perceptual dimensions can explain vowel identification by L1 and L2 talkers better than direct interpretation of L1 and L2 talkers' perception performance with a single cue-weighting function, 2) different reliable cues may be placed in different linear functions or latent variables and 3) perceptual vowel space with two latent dimensions can explain the difference in identification performance by L1 and L2 talkers.

2. Experiment: Identification test on /ɑ/, /ɔ/ and /ʌ/

2.1 Subjects

The subjects were 58 Korean university students (11 males and 47 females) who took the English Phonetics in a Korean university. The subjects' age ranges from 20 to 25 years old. All of the subjects had at least 6 years of prior English instruction at the middle and high school levels. As this experiment was conducted at the beginning of the semester, they had no phonetic knowledge of American English (hereafter, AE) vowels to be offered in the course. However, they already knew IPA symbols. None of the subjects had any reported history of speech or hearing problems.

2.2 Stimuli

2.2.1 Acoustic characteristics of the database in Hillenbrand et al. (1995)

In the current experiment, 60 vowel signals (20 for each of /ɑ/, /ɔ/ and /ʌ/) were selected from the AE vowel database in Hillenbrand et al., which are accompanied by duration, steady state F0, F1, F2 and F3 and identification rates of the signals by 20 AE talkers. In this section, it will be described how the database in Hillenbrand et al. (1995) was built up and it will be also explained how the 60 vowel signals were selected for the current experiment.

In the database, audio recordings were made of subjects reading lists containing 12 vowels: /i, ɪ, ɛ, æ, ɑ, ɔ, ʊ, u, ʌ, ɝ, eɪ, oʊ/. And subjects read from the randomized 12 hVd utterances containing the words "heed," "hid," "hayed," "head," "had," "hod," "hawed," "hoed," "hood," "who'd," "hud," "heard," "hoyed," "hide," "hewed," and "how'd." The database consisted of hVd signals of 139 AE talkers (45 men, 48 women and 46 10- to 12-year-old children (27 boys, 19 girls)). 87% of the talkers were raised in the southeastern and southwestern parts of Michigan. The rest of the talkers were from Illinois, Wisconsin, Minnesota, northern Ohio and northern Indiana. Note that an extensive screening procedure was used to select these 139 subjects from a larger group before the recording.

After the recording procedure for the database, the following measure procedures followed. The duration of the vowel part was measured by hand from high-resolution gray-scale digital spectrograms using the measurement criteria in Peterson and Lehiste (1960). A judgment of steady-state time for each signal was independently by two experimenters based on a spectral peak as shown in Figure 1 and a gray-scale spectrogram. And then 10% of the stimuli were randomly selected and remeasured for duration by the second author. The mean absolute difference between the original and remeasured duration was 6.9 ms.

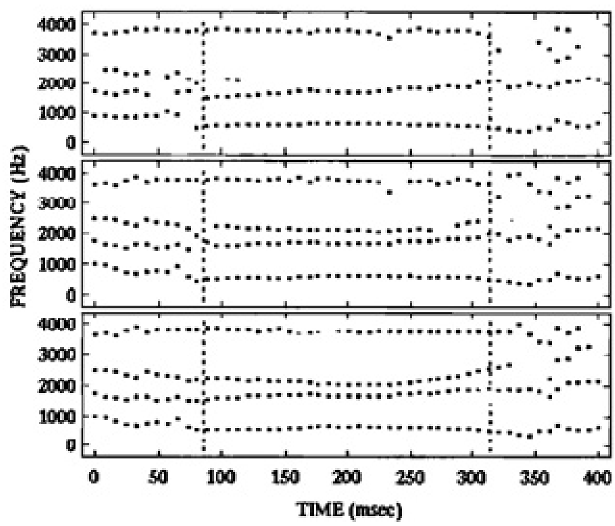

Figure 1. Spectral peak display of the word "heard" spoken by a child. The dashed vertical lines indicate the beginning and end of the vowel nucleus. The top panel shows the signal after the original 14-pole LPC analysis, the middle panel shows the signal after reanalysis with 18 poles and the bottom panel shows the signal after hand editing with a custom editing tool. (Figure 1 in Hillenbrand et al. 1995: 3100)

According to Hillenbrand et al. (1995), the formant tracks were hand edited to correct formant tracking errors through linear interpolation, based on the LPC peak display overlaid on a gray-scale spectrogram through the calculation of 14-pole and then 18-pole reanalysis of 128-point LPC spectra every 8 ms over 16 ms (256 point) Hamming windowed segments. And then the steady-state centers of F1, F2, F3 tracks were measured.

Finally, F0 contours were extracted by an autocorrelation pitch tracker in Hillenbrand (1988) and then tracking errors were corrected through the reanalyses of the signals.

2.2.2 Perceptual characteristics of the database in Hillenbrand et al. (1995)

AE talkers who conducted the identification test of the spoken signals in the database were 20 phonetically trained undergraduate and graduate students in the Speech and Pathology and Audiology Department at Western Michigan University.

65% of the vowel signals in the database were identified unanimously by the listeners and 89% were identified at rates of 90% or greater. The problem was that many talkers of general AE do not distinguish between /ɑ/ and /ɔ/ despite the fact that before the recording, the authors assessed the talkers' production of /ɑ/ and /ɔ/ in a short informal conversation with one of the authors in Hillenbrand et al. (1995) as a screen process and also the fact that the talkers were required to pass a brief discrimination test of /ɑ/ vs. /ɔ/ minimal pairs. Among the recorded signals, the intended /ɔ/ was heard as /ɔ/ 82.0%, as /ɑ/ 13.8% and as the other vowels 3.2%. On the other hand, the intended /ɑ/ was heard as /ɑ/ 92.3% and as the other vowels including /ɔ/ 7.7%. Hillenbrand et al. (1995) observed that the /ɔ/ signals consistently heard as a vowel other than that intended by the talker, were production errors which frequently occur in general American English and hence were not convincing /ɔ/ signals. They suggested that nearly 14% of the attempts at /ɔ/ were production errors despite the unsuccessful and ineffective prior dialect screening procedures which were conducted before the recording.

In order to pick up convincing /ɔ/ signals as stimuli for the present experiment, identification rates of all the vowel types in the database were compared with one another. It was found that all the vowel types except /ɔ/ were identified as intended vowels above 90%: from the highest 99.6% for /i/ to the lowest 90.8% for /ʌ/, compared to 82.0 for /ɔ/. For this reason, we assumed that the if the /ɔ/ signals are produced correctly, they should be heard as /ɔ/ by more than 90% of the AE listeners, as all the other vowels were heard as intended vowels by more than 90% of the listeners. Under this assumption, we selected the vowel signals which were produced by AE talkers whose /ɔ/ signals were agreed on the intended /ɔ/ by more than 90% of the 20 AE listeners. However, this selection process did not influence the selection of the other two vowels since all the other vowels were identified as intended vowels by more than 90% of the listeners as shown in Table 1.

Table 1. Overall correct identification by vowel category in the database (in Table VI in Hillenbrand et al. (1995)). HGCW refers to Hillenbrand et al. and PB to Peterson and Barney (1952).

	HGCW	PB
/i/	99.6	99.9
/ɪ/	98.8	92.9
/e/	98.3	*
/ɛ/	95.1	87.7
/æ/	94.1	96.5
/ɑ/	92.3	87.0
/ɔ/	82.0	92.8
/o/	99.2	*
/ʊ/	97.5	96.5
/u/	97.2	99.2
/ʌ/	90.8	92.2
/ɝ/	99.5	99.7
Total	95.4	94.4
Men	94.6	**
Women	95.6	**
Children	93.7	**

* These vowels were not recorded by Peterson and Barney.
** Peterson and barney did not report results separately for men, women, and child talkers.

2.2.3 The stimuli selected in the current experiment

Target vowels were American English /ɑ/("a"), /ɔ/("o") and /ʌ/("x") in the hVd frame as shown below[1]:

Table 2. /ɑ/, /ɔ/ and /ʌ/ in hVd

vowel	a(ɑ)	o(ɔ)	x(ʌ)
hVd	hod	hawed	hud

The selected hVd stimuli of the three vowels from the database were spoken by 8 males and 12 females. The database contains the measures of duration, F0, steady-state F1, F2 and F3 for each of the spoken hVd tokens. Each of the spoken tokens contains the accuracy results from the vowel identification test, conducted by 20 American talkers (hereafter, ATs). Table 3 shows ATs' mean identification accuracy values for the stimuli of

[1] As IPA symbols for the vowels are incompatible with the statistical program package used, /ɑ/, /ɔ/ and /ʌ/ will be transcribed as /a/, /o/ and /x/, respectively, in this paper.

the three vowels.

Table 3. Identification accuracy by ATs for /a/, /o/ and /x/

stimuli	Mean accuracy	s.d.	N
a	96.5	3.26	20
o	97.0	3.40	20
x	96.5	5.16	20
Total	96.7	3.98	60

The total number of the presentations of hVd stimuli was 60 (= (8 males + 12 females) * 3 vowels).

2.2.4 The acoustic structure of naturally spoken stimuli

In the literature on cue weighting, stimuli were constructed by editing natural signals produced by one English talker. Morrison (2002), for example, created a multidimensional continuum of perceptual stimuli to be fed in to Discriminant Analysis in order to derive cue-weighting strategies by Japanese and Spanish talkers. Based on the speech by an English talker, a multidimensional continuum was built up along the vowel spectra of F1, F2 and F3 (5 steps), vowel duration (7 steps), plosive closure duration (5 steps) and speaking rate (2 steps).

The current analysis, on the other hand, used 60 natural speech signals whose measurements of F0, F1, F2, F3 and vowel duration formed naturally created continuum as shown below:

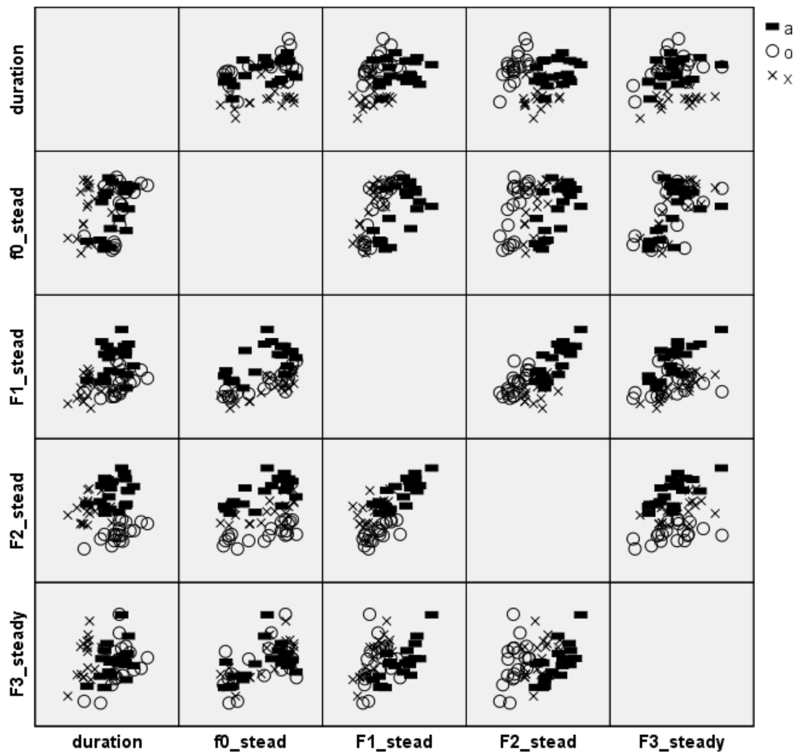

Figure 2. Scatterplot matrix of vowel duration, F0, F1, F2 and F3

The scatterplot matrix shows that the multidimensional continuum was relatively well distributed along vowel duration, F0, F1, F2 and F3. Square- or circle-shaped distribution in the matrix implies a complete two-dimensional continuum along the two cues. One potential problem is that the distribution of F1 vs. F2 forms an incomplete two-dimensional continuum. However, we will assume that this distribution is serviceable for the current purpose since the plots outside the plotted area would not be heard as /a/, /o/ and /x/ due to the acoustic nature of F1 and F2. These values of the cues will be fed to Discriminant Analysis.

The stimuli selected for the present experiment are plotted on the acoustic vowel space with F2 on the x-axis and F1 on the y-axis, as follows:

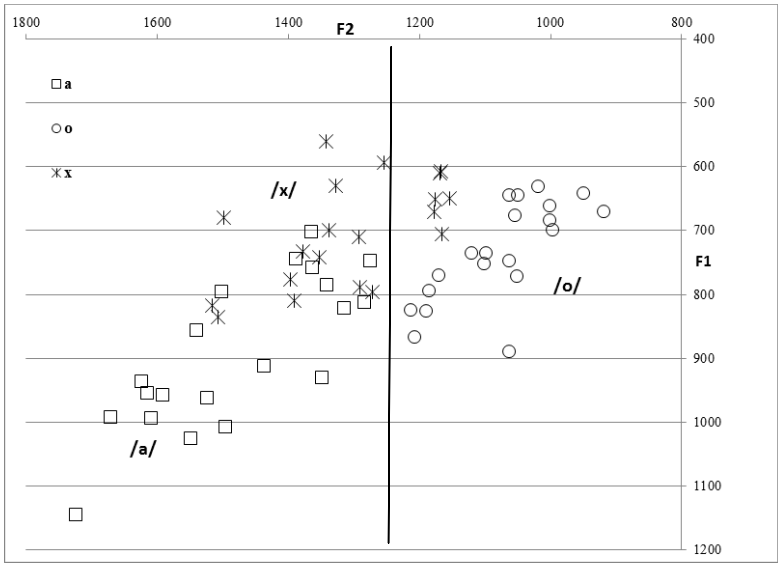

Figure 3. Acoustic vowel chart with F1 and F2 for /a/, /o/ and /x/ in the spoken stimuli

In Figure 3, the plots of the stimuli on the acoustic vowel space of F1 and F2 are widely spread across both F1 and F2 axes. There arises only one clear grouping boundary between /a/ and /o/ at 1250 Hz along F2. No further grouping boundaries are observed. This suggests that measures of F1 and F2 alone cannot distinguish between /a/ and /x/ and between /o/ and /x/. When the vowels are plotted with F0 on the x-axis and duration on the y-axis, duration might help distinguish between /a/ and /x/ plots. /x/ plots are separated from the other vowel plots at 255 ms along duration, as shown in Figure 4.

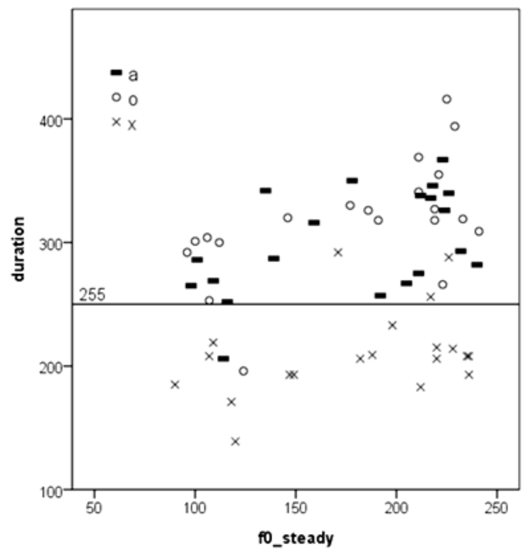

Figure 4. Plotting of stead-state F0 and duration of /a/, /o/ and /x/ of the stimuli

Then it might be hypothesized that /a/, /o/ and /x/ can be distinguishable by heavily weighting the cues like F2 and duration by ATs. And this hypothesis will be statistically verified by stepwise Discriminant Analysis (hereafter, DA), which will derive significant linear discriminant functions of differently weighted cues and builds up a perceptual vowel space with linear discriminant functions. Insignificant cue variables will be removed in the analysis through a stepwise removing procedure. As a result, two models will be built up, one for ATs and another for KTs. Then the two models will be compared with each other as to whether the two groups of talkers use the same or at least similar vowel identification strategies or not.

2.3 Procedure

Before the experiment was conducted, the IPA symbols of the three vowels were introduced to the subjects to make sure that all the Korean subjects were familiar with the symbols and actually they already were. The protocol for the experiment was set up based on Alvin 2.0 (Hillenbrand and Gayvert 2005). On the computer screen, three AE vowel icons were placed side by side with the IPA symbol of the target vowel and the hVd word containing the target vowel: *hod*, *hawed* and *hud*.

Each subject listened to each of the randomized hVd stimuli over a headphone from a PC and was forced to click on one of the 3 vowel icons: a forced-choice identification test. The sound volume was adjustable by subjects. The next stimulus was presented 500 ms after an identification click. When a subject made an error click, s/he could go back and make a readjustment click after listening to the previous presentation again. The number of replaying the presentation was limited to three times. This experiment protocol let participants proceed at their own comfortable pace, though the experiment took less than 2 minutes.

Figure 5. hVd identification test screen

3. Results

A confusion matrix of the three vowels was derived for each subject, totaling 58 confusion matrices for all the subjects. The following is the combined confusion matrix from all 58 KTs.

Table 4. Vowel identification performance by KTs

KTs' Performance	Clicked Group Membership			# of stimuli
	a	o	x	
a	874(75.3%)	235(20.3%)	51(4.4%)	1160
o	189(16.3%)	699(60.3%)	272(23.4%)	1160
x	104(9.0%)	139(12.0%)	917(79.1%)	1160
		Overall correct count (accuracy)	2490(71.6%)	3480

The total accuracy of /a/ from 58 KTs is 75.3%, /o/ 60.3% and /x/ 79.1%. The confusion matrix contains the information of correct identification rates for target vowels and of misidentification rates as each of the other two vowels. The overall accuracy for all the three vowels is 71.6%.

In order to compare KTs and ATs' identification performance across the three vowels, KTs' and ATs' identification rates of each of the 20 stimuli of each vowel type (total 60 stimuli of all the three vowel types) were recalculated to compare KTs' and ATs' identification rates for each stimulus (total 120 identification rates for all 60 stimuli). KTs' and ATs' overall performance is illustrated in the boxplots in Figure 6:

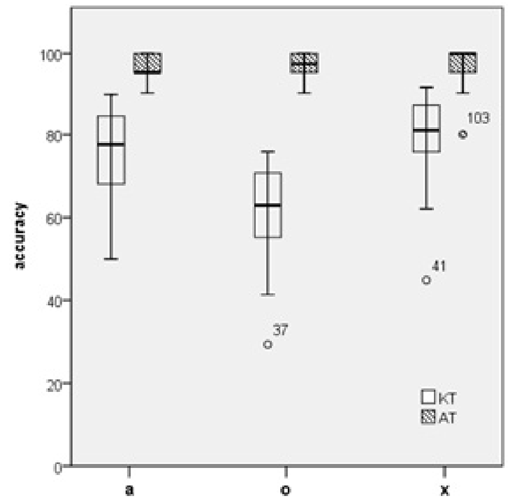

Figure 6. Accuracy boxplots for /a/, /o/ and /x/ for KTs and ATs

KTs' and ATs' performance was compared across vowel types to see if there is any identification performance difference across the three vowels (within subjects) between ATs and KTs (between groups).

Table 5. KTs' and ATs' mean vowel identification rates across 20 stimuli of each of /a/, /o/ and /x/

Vowel type	/a/	/o/	/x/	Overall mean
# of stimuli	20	20	20	60
KTs	75.2%	60.4%	79.3%	71.6%
ATs	96.5%	97.0%	96.5%	96.7%

Mixed measures ANOVA showed that there is a significant interaction effect between vowel types (/a/, /o/ and /x/) and the two groups (ATs and KTs): Mauchly's Test of Sphericity = 0.009 < 0.05, Greenhouse-Geisser $F(1.64, 62.11) = 11.15$, $p < 0.01$). This is because there is a significantly poorer performance in the identification of /o/ among the three vowels by KTs as compared to the performance in identifying the same vowels by ATs, as illustrated in Figure 6. Furthermore, There is a significant difference among with-subject vowel types due to the poor performance in /o/ by KTs: Greenhouse-Geisser $F(1.64, 62.11) = 9.95$, $p < 0.001$. There is also significant between-group difference between KTs (mean = 71.64, S.E. = 0.78) and ATs (mean = 06.67, S.E. = 0.78): $F(1, 38) = 518.93$, $p < 0.001$. This means that ATs' identification performance is far better than KTs'.

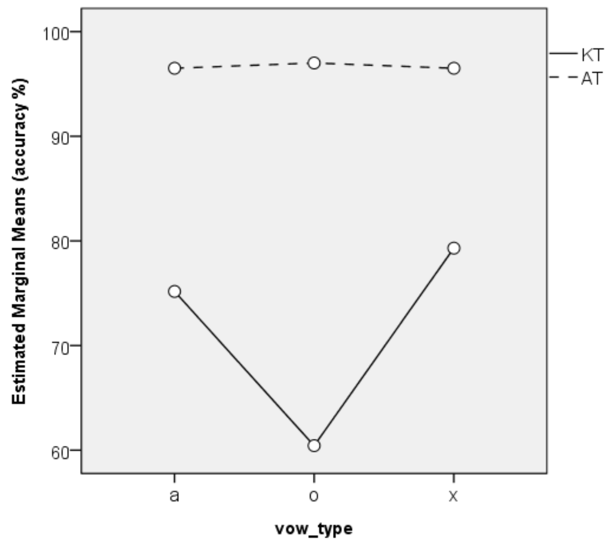

Figure 7. Estimated marginal means for /a/, /o/ and /x/ compared between KTs and ATs

4. Discussion

4.1 DA models for the performance of ATs and KTs

Stepwise DA is a multivariate analysis of variance and can test the hypothesis that English /a/, /o/ and /x/ differ significantly on linear discriminant functions of differently weighted variables (duration, F0, F1, F2 and F3). The variables within a discriminant function will have to be uncorrelated with each other. As a result, the discriminant function can explain the variance of the data as much as possible. These discriminant functions are actually latent variables in the form of linear combinations of weighted cues. They also form two separate latent dimensions for a two dimensional perceptual vowel space for English /a/, /o/ and /x/. Namely, it is possible to plot the vowel stimuli on the perceptual vowel space with the two derived latent dimensions.

The procedure of DA takes "traditionally dependent" variables like duration, F0, F1, F2 and F3) as predictor variables and subjects' vowel identification clicks as a grouping variable, to derive new latent variables, which are represented as discriminant functions of differently weighted cues. Then all the vowel stimuli and the vowel centroids are plotted on the perceptual vowel space of the discriminant functions (i.e., the derived latent variables). Note that the vowel group centroids are computed by entering variable means of the vowel group in the discriminant functions. All the vowel stimuli will be plotted together close to their vowel group centroids.

DA can also predict subjects' decision on the vowel membership of spoken vowel stimuli based on the distance between the stimulus plots and the vowel centroids computed by Euclidean Distance[2]. Alternatively, this classification can be replaced by the derived classification functions, which take the measures of duration, F0, F1, F2 and F3 of a vowel stimulus and predicts to which vowel group the vowel stimulus belong. However, these classification functions were not to be tested with new raw data in this paper, which will be attributed to further study.

From now on, separate DA models for ATs and KTs are to be derived from the identification results by ATs and KTs. And it will be shown how correctly the proposed DA models account for ATs' and KTs' identification performance.

Stepwise DA was used to conduct a multivariate analysis of variance test of the AT model in which ATs' perceptual performance between English /a/, /o/ and /x/ would be significantly represented by two derived linear functions of the four or less variables (duration, F0, F1, F2 and F3). For the two derived discriminant functions combined, the overall Chi-square test was significant (Wilks' lambda = .048, Chi-square = 168.95, df = 8, Canonical correlation = .897 and .869, p <. 001). Wilks' lambda says that combining two derived discriminant functions may account for all but 4.8% of the variance in the vowel identification performance by ATs.

In the procedure of stepwise DA, the variable that maximizes the overall Wilks' lambda at each step was entered into the analysis. And only significantly contributing variables will be kept in the final discriminant functions. Note that insignificant F3 was removed in the analysis. Two discriminant functions were derived after the stepwise procedure: Function1 = -0.012*duration – 0.008*F0 – 0.004*F1 + 0.012*F2 – 7.525 and Function2 = 0.019*duration – 0.025*F0 + 0.012*F1 + 0.001*F2 – 11.087. These two functions will be used as two latent dimensions for a perceptual vowel space.

When the structure matrix of functions and standardized variables for ATs is considered for the relative standardized contribution of the cues (namely, all cues are on same scale), the two functions are interpreted as follows. The high positive contribution of F2 (0.670) in function1 suggests that function1 can be directly interpreted as F2. On the other hand, the high positive contribution of duration (0.523) and F1 (0.466) in function2 suggests that function2 can be interpreted as "ATs use both duration and F1 cues at the same time." This means that ATs' vowel identification strategies are based on F2 on one latent dimension and duration and F1 on the other. The proposed vowel space of F2 on one dimension and F1 and duration on another diverges from the traditional acoustic vowel space of F2 on one dimension and F1 on another.

---

2 Distance((x, y), (a, b)) = $\sqrt{(x - a)^2 + (y - b)^2}$

Table 6. Structure matrix for ATs

	Func1 (Latent Dimension1)	Func2 (Latent Dimension2)
Interpretation of functions	(F2)	(Duration + F1)
F2	**.670***	.308
dur	-.327	**.523***
F1	.184	**.466***
~~F3**~~	~~.072~~	~~.223*~~
F0	.000	**-.014***
*significant **F3 is removed in the analysis after stepwise DA		

And three group centroids of /a/, /o/ and /x/ stimuli were derived by computing each vowel group's mean cue values into the functions. Each of the stimuli plots belong to the vowel group whose centroid vector the vector of the plot is closest to. The following perceptual vowel space of latent dimension1 and dimension2, shows how the two derived discriminant functions from stepwise DA discriminate between the three vowel groups based on the two functions. Note that the three filled rectangles refer to three vowel group centroids:

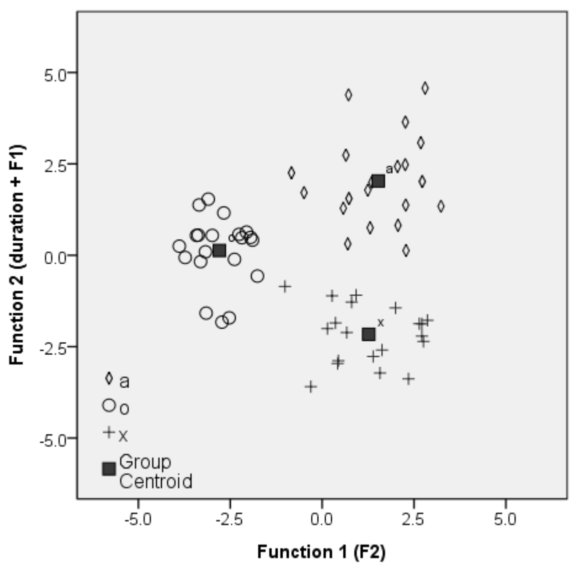

Figure 8. Vowel group centroids and stimuli plots on the perceptual vowel space of latent finction1 and function2 for ATs

In Figure 8, latent function1 is directly interpreted as F2 whereas latent function2 as the combination of differently weighted duration and F1.

The three classification functions for the model for ATs are as follows (note again that F3 was removed from the classification function): F(a) = 0.183*duration - 0.329*F0 + 0.097*F1 + 0.128*F2 - 137.419, F(o) = 0.198*duration - 0.247*F0 + 0.090*F1 + 0.076*F2 - 84.478 and F(x) = 0.105*duration - 0.220*F0 + 0.048*F1 + 0.122*F2 – 88.867.

Table 7. Classification results from AT model

count	Predicted Group Membership			# of stimuli
	a	o	x	
a	20(100%)	0	0	20
o	0	20(100%)	0	20
x	0	1(5%)	19(95%)	20
			98.3% of ATs' overall vowel clicks	60

This model for ATs classified 98.3% of ATs' identification performance with /a/ and /o/ classified 100% and /x/ classified 95%, as shown in Table 7.

Now, the model for KTs is to be built up to see if KTs use the same identification strategies for /a/, /o/ and /x/. Stepwise DA tested the variance of the model for KTs based on the same variables (duration, F0, F1, F2 and F3). For the two derived discriminant functions combined, the overall Chi-square test was significant (Wilks' lambda = .0432, Chi-square = 2919.845, df = 10, Canonical correlation = .661 and .483, p <. 001). Wilks' lambda says that combining two derived discriminant functions may account for all 56.8% of the variance in the vowel identification by KTs. The reason for the low percent of the explained variance is that KTs were not consistent in using variables identifying the vowels, which will be interpreted shortly. Even though the KT model can explain 56.8% of the variance, it can significantly show what identification strategies KTs use.

As for the interpretation of the two latent functions (see Table 8 below), the high positive contribution of F1 (0.597) in conjunction with a low negative contribution of F3 characterizes discriminant function1 as F1 with negative adjustment of F3: (F1 – F3). Therefore, this latent functional dimension may be identified more or less as F1 dimension in the traditional acoustic vowel space of F2 and F1. However, the high contribution of duration, though insignificant, is observed. On the other hand, high negative contribution of F2 (-0.707) and high positive contribution of duration (0.652) characterize latent function2 as the combination of F2 and duration (duration – F2)[3], which is more or less compared to F2 dimension in the traditional acoustic vowel space of F2 and F1. However, KTs' model

[3] F0 is ignored due to its low contribution relative to highly contributing F2 and duration in function2.

is different from the traditional acoustic vowel space in that this dimension used duration cue in addition to F2 cue.

Table 8. Structure matrix for KTs

	Func1 (Latent Dimension1)	Func2 (Latent Dimension2)
Interpretation of functions	(F1 – F3)	(Duration – F2)
F1	.597*	-.306
F3	-.111*	-.019
F2	.281	-.707*
dur	.599	.652*
F0	-.012	.135*
*significant		

Note that ATs used a combination of F1 and duration in function2 (see Table 6). In KT model, however, duration insignificantly contributed a lot (0.599) in function1. The insignificant duration coefficient strongly suggests that duration is not consistently utilized in discriminant function1 and should not be considered to make substantive contribution to function1. Namely, despite highest contribution of duration to function1, KTs' use of duration turns out to be statistically insignificant due to inconsistency. This might be one important reason why KTs identified the vowels poorly.

Note also that ATs used a combination of positive duration and positive F1 values in function2, which implies that both duration and F1 are considered in one dimension by ATs. On the other hand, KTs used a combination of positive duration values and negative F2 values in function2. It is not clear in the current analysis how to interpret this combination in the dimension, which requires further study.

The following perceptual vowel space for KTs shows how the two derived discriminant functions discriminate between the three vowel groups by plotting the vowel stimuli with the two derived latent functions. Note that the three filled rectangles refer to three vowel group centroids and a lot of overlapping plots are observed.

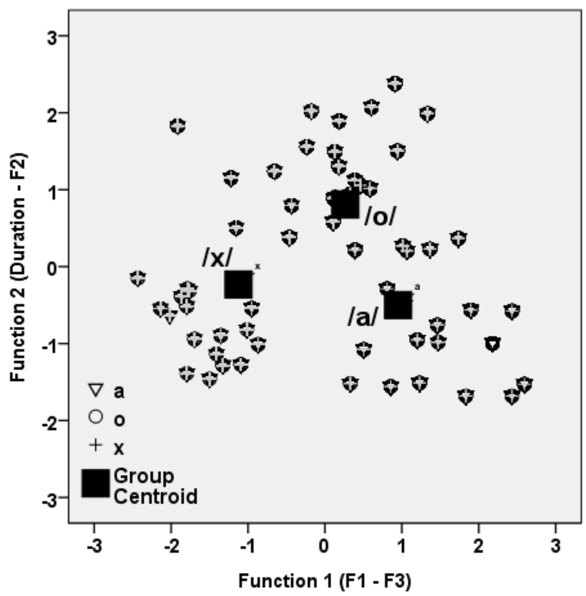

Figure 9. Vowel group centroids and stimuli plots on the perceptual vowel space of latent function1 and function2 for KTs (/a/, /o/ and /x/ plots are seriously overlapped)

The three classification functions for the model for KTs are as follows: F(a) = 0.111*duration - 0.221*F0 + 0.007*F1 + 0.038*F2 + 0.064*F3 – 113.278, F(o) = 0.118*duration – 0.202*F0 - 0.02*F1 + 0.033*F2 + 0.065*F3 – 107.280 and F(x) = 0.090*duration - 0.194*F0 - 0.012*F1 + 0.037*F2 + 0.067*F3 – 88.867.

The derived KT model can correctly classify 71.3% of overall KTs' vowel identification clicks with 71.9% of clicked /a/, 65.2% of clicked /o/ and 76.0% of clicked /x/, as shown in Table 9:

Table 9. Classification results from KT model

KT Model	Predicted Group Membership			Total Clicks by KTs
	a	o	x	
a	838(71.9%)	210(18.0%)	0117(10.0%)	1165
o	215(20%)	701(65.2%)	1590(14.8%)	1075
x	49(4%)	249(20.1%)	942(76.0%)	1240
			71.3% of KTs' overall vowel clicks	3480

4.2 AT and KT model comparison

KT and AT models simulated KTs' and ATs' identification performance on /a/, /o/ and /x/ and show that KTs and ATs used different identification strategies, which is summarized in Table 10:

Table 10. Variables of major contribution to discriminant functions for KT and AT models

	Latent Func1	Latent Func2
KTs	F1* - F3* (,**duration****)	**duration*** - F2*
ATs	F2*	**Duration*** + F1*
*significant **high contribution but insignificant due to inconsistency		

KTs' identification strategies are different from ATs' in five aspects. Firstly, KTs use duration along with F2 (in function2) whereas ATs use duration along with F1 (in function2). Second, KTs make an inconsistent use of duration along with F1 (in function1) and make a consistent use of duration along with F2 (in function2), whereas ATs make a consistent use of duration along with F1 (in function2). Third, in the AT model, individual function1 with F2 alone explained more variance (57.1% of the variance explained) than individual function2 with duration and F1 can (42.9% of the variance explained), as in Table 11:

Table 11. Variance explained by individual functions in AT model

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	4.125	57.1	57.1	.897
2	3.096	42.9	100.0	.869

In KT model, on the other hand, function1 with F1 (71.8% of the variance explained) alone explained more than function2 with F2 and duration (28.2% of the variance explained) alone, as shown in Table 12:

Table 12. Variance explained by individual functions in KT model

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.776	71.8	71.8	.661
2	.305[a]	28.2	100.0	.483

Fourth, KTs use F3 (interpreted as lip-rounding) in function1 (-.111) (Table 8) whereas ATs never use F3 in vowel identification (Table 6). These facts seem to result in KTs' overall poorer performance than ATs' and furthermore seem to result in KTs' extremely poorer performance for

/o/ than for the other two vowels. Even though /o/ is a rounded vowel, ATs did not use F3 for the perception of /o/. However, KTs tried to pick up F3 for /o/, resulting in extremely poorer identification performance for /o/. Fifth, whereas ATs use highly contributing F2 in function1 and highly contributing duration and F1 together in function2, KTs use F1 and F3 together for function1 and F2, duration for function2, as in Table 10. KTs were picking up duration at a chance level for function1 which should have been significantly used together with F1 within the same function and were picking up an unreliable F3 instead, as shown in Table 10. As for function2, they picked up unreliable duration and F0. This means that KTs were being confused as to which cues are significantly contributing to what extent in each function.

On the other hand, KTs and ATs' models share the following strategies: 1) F1 and F2 are separated into two different functions, 2) KTs used duration, though quiet inconsistently and insignificantly, along with F1 in function1. This strategy, if duration were significantly used, would resemble ATs' strategies in function1. This suggests that KTs' relatively poor performance resulted critically from the (un)critical use of duration.

As was observed, KTs considered duration along with F2 while ATs use duration along with F1. Then it can be suggested that if they considered duration along with F1 only in vowel identification, as shown in Table 13, their performance would go up drastically, though this is a pending question for further study on vowel training:

Table 13. Variables of major contribution to discriminant functions for KT and AT models: Duration switch across functions

	Latent Func1	Latent Func2
KTs	(F1* - F3*) + duration**	(**duration*** - F2)
ATs	(F2*)	(Duration* + F1*)
*significant **high contribution but insignificant		

Furthermore, in AT Model, as shown in Table 11, function1 with F2 alone can explain 57.1% of the variance whereas function2 with duration and F1 alone 42.9%. In KT model, however, function1 with F1 (71.8% of the variance explained) explained more than individual function2 with F2 and duration (28.2% of the variance explained), as shown in Table 12. Therefore, for a better performance by KTs, function2 would have to explain more variance than function1. Namely, the role of function2 would have to be raised to more than the level of that of function1.

5. Conclusion

In this paper, we tried to model ATs' and KTs' perception performance on English /ɑ/, /ɔ/ and /ʌ/, using stepwise Discriminant Analysis. And it was proposed that ATs and KTs may use two latent variables (or perceptual dimensions) which are identified as two differently weighted cue functions. In previous studies on cue weighting, L1/L2 talkers' perception performance on the binary distinction between /i/ and /ɪ/ was approached with a single function with differently weighted cues. In the present paper, however, two significant latent variables were derived as functions of differently weighted cues. They together plot centroids of the vowel groups and the stimulus on the perceptual vowel space. And the distance between the plotted stimulus and the vowel centroids was calculated. As a result, the plotted stimulus will be classified in the vowel group the centroid of which the plotted stimulus is closest to. The present analysis shows that KTs and ATs use different identification strategies for the perception of English /ɑ/, /ɔ/ and /ʌ/, by building two separate DA models for KTs and ATs.

This paper, however, did not address several important questions. First, the current analysis did not divide KTs into subgroups according to their vowel identification performance. As a result, it is not clear whether KTs might show different identification strategies depending on their English vowel perception performance levels. However, it might be possible that the identification strategies which KTs of high performance use may resemble those of ATs' whereas KTs of poor performance may use different identification strategies. Second, the models for KTs and ATs were not tested in this study with new data. Third, the current study was limited to the identification of three English vowels only. If all the English monophthongs were tested, it would be possible for the number of latent dimensions to increase to three and as a result, a three-dimensional vowel space might have to be built up for an optimal model. Then interpretation problems of the resulting three-dimensional perceptual vowel space might arise. Fourth, L2 vowel perception training may change L2 talkers' perception performance (Nishi and Kewley-Port 2007, 2008). Then will vowel perception training move a critical cue (for example, duration) in one latent variable to the other latent variable, resembling the way ATs use weighted cues for vowel perception (Ylinen 2009)? These are pending questions for further research.

Appendix. List of vowel stimuli selected from the database in Hillenbrand et al. (1995)

File name	Vowel type	ID rate by ATs	Duration	F0	F1	F2	F3
m01ah.wav	/ɑ/	95	316	159	813	1283	2687
m44ah.wav	/ɑ/	95	286	101	822	1315	2708
w03ah.wav	/ɑ/	95	257	192	1025	1548	2748
w09ah.wav	/ɑ/	90	282	240	913	1436	2589
w10ah.wav	/ɑ/	95	367	223	856	1540	2667
w14ah.wav	/ɑ/	100	275	211	994	1609	2930
w15ah.wav	/ɑ/	100	340	226	955	1615	2678
w17ah.wav	/ɑ/	95	267	205	937	1623	2844
w22ah.wav	/ɑ/	95	350	178	1008	1495	3018
w30ah.wav	/ɑ/	100	293	232	957	1591	2733
w38ah.wav	/ɑ/	100	336	217	993	1671	2751
m07ah.wav	/ɑ/	100	265	98	786	1341	2403
w39ah.wav	/ɑ/	95	338	212	931	1348	2698
w48ah.wav	/ɑ/	95	346	218	796	1502	2810
w49ah.wav	/ɑ/	100	326	224	1145	1723	3272
m18ah.wav	/ɑ/	95	342	135	702	1364	2498
m21ah.wav	/ɑ/	95	206	114	758	1363	2421
m35ah.wav	/ɑ/	90	269	109	748	1274	2406
m36ah.wav	/ɑ/	100	252	116	744	1388	2541
m39ah.wav	/ɑ/	100	287	139	963	1524	2552
m01aw.wav	/ɔ/	95	330	177	699	996	2548
m44aw.wav	/ɔ/	100	301	100	684	1001	2833
w03aw.wav	/ɔ/	100	326	186	772	1051	2870
w09aw.wav	/ɔ/	95	309	241	890	1063	2523
w10aw.wav	/ɔ/	100	416	225	771	1170	2760
w14aw.wav	/ɔ/	95	266	223	748	1063	2876
w15aw.wav	/ɔ/	100	394	229	827	1189	2633
w17aw.wav	/ɔ/	95	341	211	794	1185	2896
w22aw.wav	/ɔ/	100	318	191	736	1120	3055

w30aw.wav	/ɔ/	95	319	233	867	1207	2746
w38aw.wav	/ɔ/	90	327	219	825	1213	2751
m07aw.wav	/ɔ/	90	292	96	642	949	2437
w39aw.wav	/ɔ/	100	369	211	752	1101	2616
w48aw.wav	/ɔ/	100	355	221	736	1098	2821
w49aw.wav	/ɔ/	100	318	219	677	1054	3278
m18aw.wav	/ɔ/	100	320	146	662	1001	2744
m21aw.wav	/ɔ/	95	253	107	631	1019	2219
m35aw.wav	/ɔ/	95	304	106	645	1049	2507
m36aw.wav	/ɔ/	95	300	112	645	1063	2565
m39aw.wav	/ɔ/	100	196	124	670	919	2241
m01uh.wav	/ʌ/	95	292	171	670	1178	2566
m44uh.wav	/ʌ/	95	208	107	650	1154	2640
w03uh.wav	/ʌ/	80	206	182	789	1291	2877
w09uh.wav	/ʌ/	95	193	236	809	1391	2927
w10uh.wav	/ʌ/	90	256	217	680	1499	2885
w14uh.wav	/ʌ/	100	206	220	742	1352	2960
w15uh.wav	/ʌ/	95	288	226	700	1338	2750
w17uh.wav	/ʌ/	95	183	212	705	1166	2890
w22uh.wav	/ʌ/	100	209	188	777	1397	3033
w30uh.wav	/ʌ/	100	208	235	835	1507	2834
w38uh.wav	/ʌ/	100	214	228	818	1516	2970
w39uh.wav	/ʌ/	100	233	198	710	1293	2623
w48uh.wav	/ʌ/	100	208	236	733	1378	2891
w49uh.wav	/ʌ/	100	215	220	796	1271	3197
m07uh.wav	/ʌ/	95	185	90	651	1176	2564
m18uh.wav	/ʌ/	100	193	147	610	1168	2529
m21uh.wav	/ʌ/	90	139	120	593	1254	2302
m35uh.wav	/ʌ/	100	219	109	561	1343	2498
m36uh.wav	/ʌ/	100	171	118	630	1328	2554
m39uh.wav	/ʌ/	100	193	149	608	1167	2523

REFERENCES

BOERSMA, PAUL and PAOLA ESCUDERO. 2005. Measuring relative cue weighting: A reply to Morrison. *Studies in Second Language Acquisition* 27.4, 607-617.

ESCUDERO, PAOLA and PAUL BOERSMA. 2004. Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26.4, 551-585.

FLEGE, JAMES E. 1988. The production and perception of speech sounds in a foreign language. In H. Winitz (ed.). *Human Communication and Its Disorders, a Review*, 224-401. Norwood, NJ: Ablex.

FLEGE, JAMES E. and JAMES M. HILLENBRAND. 1986. Differential use of temporal cues to the /s/–/z/ contrast by native and non-native speakers of English. *JASA* 79.2, 508-517.

HILLENBRAND, JAMES M. 1988. MPITCH: An autocorrelation fundamental-frequency tracker. [Computer Program], Western Michigan University, Kalamazoo, MI.

HILLENBRAND, JAMES M., MICHAEL J. CLARK and ROBERT A. HOUDE. 2000. Some effects of duration on vowel recognition. *JASA* 108.6, 3013-3022.

HILLENBRAND, JAMES M. and ROBERT GAYVERT. 2005. Open source software for experiment design and control. *Journal of Speech, Language and Hearing Research* 48.1, 45-60.

HILLENBRAND, JAMES M., LAURA A. GETTY, MICHAEL J. CLARK and KIMBERLEE WHEELER. 1995. Acoustic characteristics of American English vowels. *JASA* 97.5, 3099-3111.

HOLT, LORI L. and ANDREW J. LOTTO 2006. Cue weighting in auditory categorization: Implications for first and second language acquisition. *JASA* 119.5, 3059-3071.

HONG, SOONHYUN. 2011. The visual representation of perception of American English vowels by university-level Korean listeners: A multidimensional scaling-based model. *Studies in Phonetics, Phonology and Morphology* 17.1, 121-144. The Phonology-Morphology Circle of Korea.

________________ 2012. The relative perceptual easiness between perceptually assimilated vowels for university-level Korean learners of American English and measurement bias in an identification test. *Studies in Phonetics, Phonology and Morphology* 18.3, 491-511. The Phonology-Morphology Circle of Korea.

INGRAM, JOHN C. and SEE-GYOON PARK. 1997. Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics* 25, 343-370.

IVERSON, PAUL, PATRICIA KUHL, REIKO R. AKAHANE-YAMADA, EUGEN DIESCH, YOH'ICH TOHKURA ANDREAS KETTERMANN and CLAUDIA SIEBERT. 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87, 47-57.

MORRISON, GEOFFREY S. 2002. Perception of English /i/ and /ɪ/ by Japanese and Spanish Listeners: Longitudinal results. In Geoffrey S. Morrison and Les Zsoldos (eds.). *Proceedings of the North West Linguistics Conference* 2002, 29-48. Simon Fraser University.

NISHI, KANAE and DIANE KEWLEY-PORT. 2007. Training Japanese Listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language and Hearing Research* 50, 1496-1509.

________________________________ 2008. Nonnative Speech Perception Training Using Vowel Subsets: Effects of Vowels in Sets and Order of Training. *Journal of Speech, Language and Hearing Research* 51, 1480-1493.

PETERSON, GORDEN E. and HAROLD L. BARNEY. 1952. Control methods used in a study of the vowels. *JASA* 24, 175-184.

PETERSON, GORDEN E. and Ilse LEHISTE. 1960. Duration of syllable nuclei in English. *JASA* 32.6, 693-703.

YANG, BYUNGGON. 2010. College students' production and perception of English vowels. *English Language Teaching* 22.4, 165-184.

YLINEN, SARI, MARIA UTHER, ANTTI LATVALA, SARA VEPSÄLÄINEN, PAUL IVERSON, REIKO AKAHANE-YAMADA and RISTO NÄÄTÄNEN. 2009. Training the brain to weight speech cues differently: A study of Finnish second-language users of English. *Journal of Cognitive Neuroscience* 22.6, 1319-1332.

Soonhyun Hong
Department of English Language and Literature
Inha University
253 Yonghyun-dong, Nam-gu, Incheon
Korea 402-751
e-mail: shong@inha.ac.kr

received : October 14, 2013
revised : November 30, 2013
accepted : December 6, 2013