1998

The perception of vowels in hearing and hearing-impaired subjects

Nicole Wyzinski

Recommended Citation

Wyzinski, Nicole, "The perception of vowels in hearing and hearing-impaired subjects" (1998). *Independent Studies and Capstones.* Paper 520. Program in Audiology and Communication Sciences, Washington University School of Medicine. http://digitalcommons.wustl.edu/pacs_capstones/520

✓

The Perception of Vowels in Hearing and Hearing-Impaired Subjects

By: Nicole N. Wyzinski
Supervised By: James D. Miller, Ph.D.

Spring 1998

INTRODUCTION

Researchers, such as Nabelek, Dorman, and Leek have demonstrated that noise affects the perception of consonants and vowels. Since it is well known that consonants carry 95% of speech intelligibility, any reduction in the perception of consonants will reduce an individual's ability to perceive what was spoken. Theoretically, if measures were taken to improve the perception of consonants, speech intelligibility should increase and it does; however, accurately providing consonant cues is next to impossible without altering some characteristics of the consonant: we cannot change the place or manner of production.

Since vowels carry only 5% of speech intelligibility, they have been disregarded as unimportant components in speech perception, yet this may not be the case. While the energy of consonants is easily masked by the energy of noise, vowels are harder to mask. In addition, vowel transitions give cues to help identify consonants. Lastly, formant information is still available in environments with poor signal-to-noise ratios. Dubno and Dorman addressed this issue in 1987 when they found that individuals were able to recognize vowels in noise even though the spectral peaks and valleys were reduced.

For normal hearing listeners to correctly identify front vowels, only the first formant needs to be well specified (Dubno & Dorman, 1987). So, in noise, where a broad region of energy masks much of the spectral information, normal hearing listeners have enough of a cue to still identify the vowel. This good identification remains until the first formant is widened to six times its normal value.

Leek and Dorman (Leek & Dorman, 1987) reasoned that the frequencies of the spectral peaks signal vowel identity, so that the peak and trough amplitude differences must be maintained to some degree in the internal auditory representation when cochlear function is normal. When cochlear function is abnormal, research suggests that the amplitude differences between peaks and troughs, the peak-to-valley ratio, are reduced in the internal auditory representation. This effect is believed to be related to a widened bandwidth that can occur as a result of cochlear damage or noise or a combination of both.

As formant bandwidth broadens, speech intelligibility is effected (Summerfield & Sidwell, 1985; VanVeen & Houtgast, 1985; Boers, 1980). This effect is considered to be due to a relationship between bandwidth and peak-to-valley ratios of vowels. Broad bandwidths reflect smaller peak-to-valley ratios and narrow bandwidths reflect larger peak-to-valley ratios. In theory, then, smaller peak-to-valley ratios should have an adverse effect on speech perception and larger peak-to-valley ratios should enhance speech perception.

Improved vowel perception, then, should be considered as a valuable means of improving speech intelligibility as a whole.

Consequently, this study is being conducted to ultimately establish a link between noise, peak-to-valley ratios, and vowel perception. In addition, this study looks at the relationship between spectral resolution and vowel perception to determine if vowel perception indeed is improved and an individual's vowel space is increased, does an individual achieve better understanding of all speech?

STIMULI

Spectral Envelope Synthesizer

Four vowels, /i/, /ae/, /a/, and /u/, were synthesized using the spectral envelope synthesizer of Miller, Heidbreder, and Lee, 1996. The advantage of their spectral synthesizer is that it allows one to control individual characteristics of the spectral envelopes of speech sounds. So, in this case, we were able to vary the peak-to-valley ratio of each vowel, while still maintaining other spectral information.

In order to utilize the spectral envelope synthesizer, the center frequencies of the locations of five spectral peaks must be specified. For this study, the center frequencies of the first three formants were set equal to the mean values reported by Peterson and Barney (1952) for male talkers. The center frequencies of the fourth and fifth formants were fixed at 3500 Hz and 4500 Hz. The durations were set to 180 msec and the amplitudes were feathered with a one-half cosine multiplier with a rise time of 15 msec and a fall time of 45 msec.

The nominal peak-to-valley depths chosen were 99 dB, 36 dB, 18 dB, and 9 dB. The actual peak-to-valley depths were adjusted in accordance with the logarithmic distances between formant center frequencies such that they were increased slightly for large distances and decreased slightly for short distances as this mimics actual speech.

Loudness Balancing

Once the stimuli were synthesized, it was noted that the vowels could be differentiated on the basis of loudness differences. To counter this, a LabView program was written by Heidbreder wherein a subject judged the loudness of a test vowel against a standard vowel. The subject had to identify when the test vowel was louder or softer than the standard. The test vowels consisted of the four synthesized vowels from the spectral envelope synthesizer that varied in peak-to-valley depths. The standard stimulus was a neutral vowel.

The intensity levels at which each subject chose as sounding of the same "loudness" as the neutral vowel were averaged across subjects and a standard set of vowels balanced for "loudness" was created.

Methods

The SubGame

The "Submarine Game" is based on the military's IFF (Identification Friend or Foe) procedures. The player is a submarine commander. When a ship appears on the screen, traveling from left to right, the commander can trigger a "sonar probe" that is sent from the submarine out to the ship. If the ship returns or "echoes" the same sound, the player must recognize the sameness of the two sounds and classify the ship as a Friend. If the ship returns or "echoes" a different sound than the probe, then the commander must recognize the difference and classify the ship as an enemy or Foe. Now the player or "commander" can also fire a "torpedo" at the ship. If the player fires on a Friend, he/she loses points. If the player torpedoes a Foe, the score depends on how quickly it is done. If the player fires quickly on a Foe, he/she gets credit for a sinking. If he/she fires with a moderate delay, the player gets credit for "damaging" the Foe. If he/she is very slow to respond to a Foe, the event is scored as a "miss". If the player fails to fire on a Foe, then the enemy ship drops a "depth charge" and the submarine is "lost" and points are deducted.

Subjects initially practiced listening to the synthesized stimuli using an Identification Task. All subjects listened in soundfield via a loudspeaker. The

Identification Task is a LabView program created by Heidbreder and Miller in which the subject listens to stimuli and names which sound they hear by pressing the appropriate key. To help with identification, each stimulus is associated with a key word that is displayed on the computer screen. For /i/ the word displayed is "beet". For /ae/ the word displayed is "bat". For /a/ the word displayed is "father". For /u/ the word displayed is "boot".

After a subject has familiarized with the vowels through doing the ID Task, the actual experimental runs are initiated.

An experimental run consists of 25-trial ID tasks followed by a "Demo Run" of the SubGame, which consists of 32 critical trials. Each run uses only one peak-to-valley depth. Therefore four runs are needed to complete one experimental cycle and four such cycles were programmed using random Latin Squares. During all trials, whether ID or SubGame, the vowels were randomly varied in intensity so that a stimulus could be presented at an intensity level from -5 dB, -2.5 dB, 0 dB, 2.5 dB, or 5 dB re 74 dB SPL. This helped to ensure randomization of the stimuli so that no cues aside from peak-to-valley depth were available. Therefore, individuals' results could be attributed to formant locations and peak-to-valley ratio differences and not subjective "loudness" differences in the stimuli.

Subjects

Five subjects participated in this experiment. Two subjects were normally hearing and three subjects were hearing-impaired. Of the hearing-impaired subjects, two were hearing-aid users and one was a cochlear implant user. "Normal" hearing subjects demonstrated through a hearing screening performed at Central Institute for the Deaf pure-tone thresholds of better than 25dB HL at 500 through 4000Hz. Hearing-impaired subjects wearing hearing aids demonstrated through a hearing screening a bilateral sensorineural hearing loss of moderate-to-severe degree. Hearing-impaired subjects wearing a cochlear implant demonstrated a bilateral profound sensorineural hearing loss.

For purposes of confidentiality, the hearing-impaired hearing aid users were labeled HA1 and HA2, the cochlear implant user was labeled CI1, and the normal hearing subjects were labeled NHA1 and NHA2.

HA1 was an 86 year old male with a bilateral mild sloping to severe sensorineural hearing loss. HA2 was a 61 year old female with a bilateral moderate sensorineural hearing loss. CI1 was a 51 year old female with bilateral profound sensorineural hearing loss. NHA1 was a sixty-seven year old male with hearing within normal limits. NHA2 was a twenty-four year old female with hearing within normal limits.

ANALYSIS

Multidimensional Scaling

The results from the trials were analyzed using multidimensional scaling (mds). Subjects judged whether two stimuli were the same or different responding as quickly as possible or else points were lost. So, the subject could not simply state that a difference existed. The subject had to respond as quickly as possible.

The latency of response was used as a measure of stimulus similarity. If the subject had a long latency of response the stimuli were considered to be similar. If the subject had a very short latency the stimuli were considered to be dissimilar.

Mds used the latencies expressed as the log to the base 10 of the latency measured in milliseconds to arrange the stimuli in a perceptual space. The newly derived spaces were then compared with the familiar vowel quadrilateral.

RESULTS

Latencies

We calculated the median of the log latencies for each of the six pairs of "different" stimuli: that is; /i/ versus /ae/; /i/ versus /a/; /i/ versus /u/; /ae/ versus /a/; /ae/

versus /u/; and /a/ versus /u/. To characterize a particular talker or valley depth, we took the mean of the relevant median log latencies.

As the valley depth decreased from 99dB to 36dB to 18dB to 9dB, latency values grew larger. CI1's log latencies increased from 2.797 at 99dB to 2.746 at 36dB to 3.003 at 18dB to 3.109 at 09dB. HA1's log latencies increased from 2.615 at 99dB to 2.687 at 36dB to 2.732 at 18dB to 2.858 at 09dB. HA2's log latencies increased from 2.399 at 99dB to 2.474 at 36dB to 2.460 at 18dB. NH1'a log latencies increased from 2.471 at 99dB to 2.530 at 36dB to 2.549 at 18dB to 2.872 at 09dB. NH2's log latencies increased from 2.337 at 99dB to 2.349 at 36dB to 2.395 at 18dB to 2.646 at 09dB. This occurred across hearing and hearing-impaired individuals, hearing aid and cochlear implant users. Standard deviations from the mean (SD) were variable across subjects and valley depths. They ranged from 0 to .312 log units with a typical value near .182 log units.

The observed increases in latencies with decreasing valley depths imply that the vowels become more and more perceptually similar as the valley depth increased. In the description of the perceptual maps for vowels that follows, all spaces have been normalized by the multidimensional scaling (mds) programs to have similar maximum distances between vowels. Thus, the rather dramatic decrease in intervowel distances and areas of the vowel maps implied by the latency changes across conditions does not appear.

Rather, it is changes in vowel map shapes that are apparent in the next section.

Vowel Maps

Introduction: The vowel maps obtained from the multiple dimensional scaling program were all obtained using "the Systat" software package. The number of possible dimensions was set to two (2); the Kruskal stress algorithm was used; and linear regression between model distance and log latency was assumed. The resulting vowel maps were "normalized" so that maximum distances were similar for all conditions. The maps were plotted such that the dimension with the most "separation" is Dimension I and is on the horizontal axis.

For ease of comparison, all spaces were rotated such that a line connecting /i/ and /u/ was always paralleled to the horizontal axis, Dimension I, and such that the /ae/, /a/, or both fell below /i/ and /u/ along Dimension II or the vertical axis.

Most of the time, spaces reminiscent of the familiar phonetic "vowel equilateral" were produced. However, some times /ae/ and /a/ were "interchanged" with /ae/ being closer to /u/ than to /i/, which would not be ordinarily expected. This interchange of /ae/ and /a/ in the vowel space is not fully understood at this time and may be related to small errors in measuring their similarities with /i/ and /u/. Other pathologies included "squished" spaces with /ae/ and /a/ close to each other, or with /ae/, /a/, and /u/ all close together, or with little separation on Dimension II. The spaces so obtained for valley depths of 99, 36, 18, and 09dB, will now be described and discussed. (See Figures 1-25 in Appendix A)

Relation to Valley Depths

NHA1: For this normally hearing listener, the vowel spaces for valley depths of 99 and 36dB appear similar to the traditional vowel spaces. Note, however, that the spaces become distinctly reduced at valley depths of 18 and 09dB especially along Dimension II, which appears to correlate with "vowel height." At 18dB, note the reversal of positions for /ae/ and /a/.

NHA2: This normally hearing listener shows vowel spaces at all valley depths that are similar to the traditional vowel equilateral. The reversal of /ae/ and /a/ for the 99dB condition was probably an error resulting from small errors in measuring the similarities of /ae/ and /a/ with /i/ and /u/.

HA2: This hearing aid user was only tested with valley depths of 18, 36, and 99dB. All three of her spaces were similar to the traditional "vowel equilateral."

HA1: This hearing aid user seemed to demonstrate good separations among the vowels but their locations often appear to be idiosyncratic and somewhat unpredictable in terms of traditional vowel theory.

CI1: The cochlear implant user demonstrated good separations among the vowels for valley depths of 99 and 36dB and distinct reduction in separations for depths of 18 and 09dB. For this listener the positions of /ae/ and /a/ were reversed for all four conditions as compared to the traditional vowel equilateral.

Averages and Their Relations to a Formant Space

Maps obtained by averaging each listener's latencies across valley depths will now be described.

Auditory Perceptual Space (APS)

The APS of Miller (Miller, 1989) is a vowel spaced on formant and fundamental relations. The distances in their space were submitted to the same mds analysis, as were the log latencies of our listeners. The resulting space is shown in Figure 20. It has the appearance of the traditional vowel quadrilateral.

The average spaces obtained from the listeners can now be compared to each other and to the mds version of Miller's APS.

It can be seen that the average spaces for NHA1, for NHA2, and HA2 are very similar to the formant based APS. Indeed, the correlations between the intervowel distances between APS and each individual listener are good ranging from .81 to .94.

The spaces based on average latencies for HA1 and CI1 appear to be abnormal. For HA1 not only are /ae/ and /a/ interchanged, Dimension II, vowel height is clearly reduced. For CI1, the locations of /ae/ and /a/ are also interchanged and Dimension I appears to be slightly reduced. Nevertheless, the intercorrelations of distances with distance in APS remain good if slightly reduced being .86 for HA1 and .81 for CI1.

DISCUSSION

Just as expected, as the valley depth decreased, latencies grew larger. It took subjects more time between hearing the sonar probes to discriminate and respond as similar or different. With less perceptual information, they required additional time to process the signal. Subjects often judged two different stimuli as similar, unable to perceive the difference. With more perceptual information, subjects expressed confidence in their ability to discriminate between stimuli and thus responded quicker. They were able to identify that a difference, dissimilarity, existed between the stimuli. This relationship was seen across all subjects, hearing versus hearing-impaired, hearing aid versus cochlear implant. No age or gender effects were observed.

Hearing individuals, on average, had better vowel maps at all peak-to-valley depths than hearing-impaired individuals. The three most common errors include the inability to differentiate /ae/ and /a/, an /ae/, /a/ reversal, and an /a/, /u/ reversal. These errors, as stated earlier, are not fully understood and may be related to small errors in measuring their similarities.

While some errors did occur, intervowel distances were still maintained. In addition, the subjective space of all subjects was highly correlated with the objective APS of Miller. Using a linear analysis, a stronger correlation was noted between the hearing-impaired individuals and NHA1 than NHA2. This difference was not found when a nonlinear analysis was utilized.

Looking at all the results, two significant findings occurred. First, HA2's accurate vowel maps, good vowel space size, and high correlation with Miller's APS, was significantly better than HA1 and CI1. Second, NHA2 was better able to separate /ae/ and /a/, low vowels, from /u/. Comparing subjects, it was noted that NHA2 and NHA1 possessed excellent speech discrimination scores, followed by HA2's good speech discrimination scores, and HA1 and CI1's poor scores. These scores agree with NHA1, NHA2, and HA2's better vowel maps and higher correlation values as opposed to HA1 and CI1 pointing to a possible link between speech discrimination ability, vowel map accuracy and vowel map size.

As hypothesized in this study, vowel map space did improve as valley depths increased across all hearing and hearing-impaired subjects. Whether increasing an individual's vowel space will improve vowel transitions and consonant perception requires further investigation. The results from this study, though, show promise for substantiating such a hypothesis.

Future studies need to include more subjects, more vowel sounds, other measures for validation, and a method needs to be devised to adjust the mds results to reflect an absolute reference to clarity or correctness of perception.

Acknowledgements

I would like to thank Dr. James D. Miller for his supervision of and guidance with my research project. I would also like to thank Arnold Heidbreder for all his computer expertise and technical assistance in addition to Dr. Manabe and Dr. Dooling for making modifications to the SubGame as Dr. Miller and I requested.

References

Boers, P.M. (1980). Formant enhancement of speech for listeners with sensorineural hearing loss. *IPO Annual Progress Report*, 15, 21-28.

Dubno, J.R., and Dorman, M.F. (1987). Effects of spectral flattening on vowel identification. *Journal of the Acoustical Society of America*, 82(5), 1503-1510.

Fahey, R.P., and Diehl, R.L. (1996). Perception of back vowels: Effects of varying F1-F0 Bark distance. *Journal of the Acoustical Society of America*, 99(4), 2350-2357.

Leek, M.R., Dorman, M.F., and Summerfield, Q. (1987). Minimum spectral contrast for vowel identification by normal-hearing and hearing -impaired listeners. *Journal of the Acoustical Society of America*, 81(1), 148-154.

Miller, J.D. (1989). Auditory-perceptual interpretation of the vowel. *Journal of the Acoustical Society of America*, 85(5), 2114-2134.

Moore, B.C.J., and Alcantara, J.I. (1996). Vowel identification based on amplitude modulation. *Journal of the Acoustical Society of America*, 99(4), 2332-2343.

Nabelek, A.K., and Letowski, T.R. (1985). Vowel confusions of hearing-impaired listeners under reverberant and nonreverberant conditions. *Journal of Speech and Hearing Disorders*, 50, 126-131.

Nabelek, A.K., and Dagenais, P.A. (1986). Vowel errors in noise and in reverberation by hearing-impaired listeners. *Journal of the Acoustical Society of America*, 80(3), 741-748.

Nabelek, A.K. (1988). Identification of vowels in quiet, noise, and reverberation: Relationships with age and hearing loss. *Journal of the Acoustical Society of America*, 84(2), 476-484.

Nabelek, A.K., Czyzewski, Z., and Crowley, H. (1994). Cues for perception of the diphthong /ai/ in either noise or reverberation. Part I. Duration of the transition. *Journal of the Acoustical Society of America*, 95(5), 2681-2693.

Owens, E., Talbott, D., and Schubert, E. (1968). Vowel discrimination of hearing-impaired listeners. *Journal of Speech and Hearing Research,* 11, 648-655.

Peterson, G.E., and Barney, H.L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America,* 24, 175-184.

Pickett, J.M. (1957). Perception of vowels heard in noises of various spectra. *Journal of the Acoustical Society of America,* 29, 613-620.

Schubert, E.D. (1968). Vowel discrimination of hearing-impaired listeners. *Journal of Speech and Hearing Research*, 11, 648-655.

Sidwell, A., and Summerfield, Q. (1985). The effect of enhanced spectral contrast on the internal representation of vowel-shaped noise. *Journal of the Acoustical Society of America*, 78, 495-506.

ter Keurs, M., Festen, H.M., and Plomp, R. (1992). Effect of spectral envelope smearing on speech reception. *Journal of the Acoustical Society of America*, 91(5), 2872-2880.

Tong, Y.C., Lim, H.H., and Clark, G.M. (1988). Synthetic vowel studies on cochlear implant patients. *Journal of the Acoustical Society of America*, 84(3), 876-887.

VanTasell, D/J/. Fabry. D.A., and Thibodeau, L.M. (1987). Vowel identification and vowel masking patterns of hearing-impaired subjects. *Journal of the Acoustical Society of America*, 81, 1586-1597.

VanVeen, T.M., and Houtgast, T. (1985). Spectral sharpness and vowel dissimilarity. *Journal of the Acoustical Society of America*, 77, 628-634.

APPENIDX A

FIGURE 1

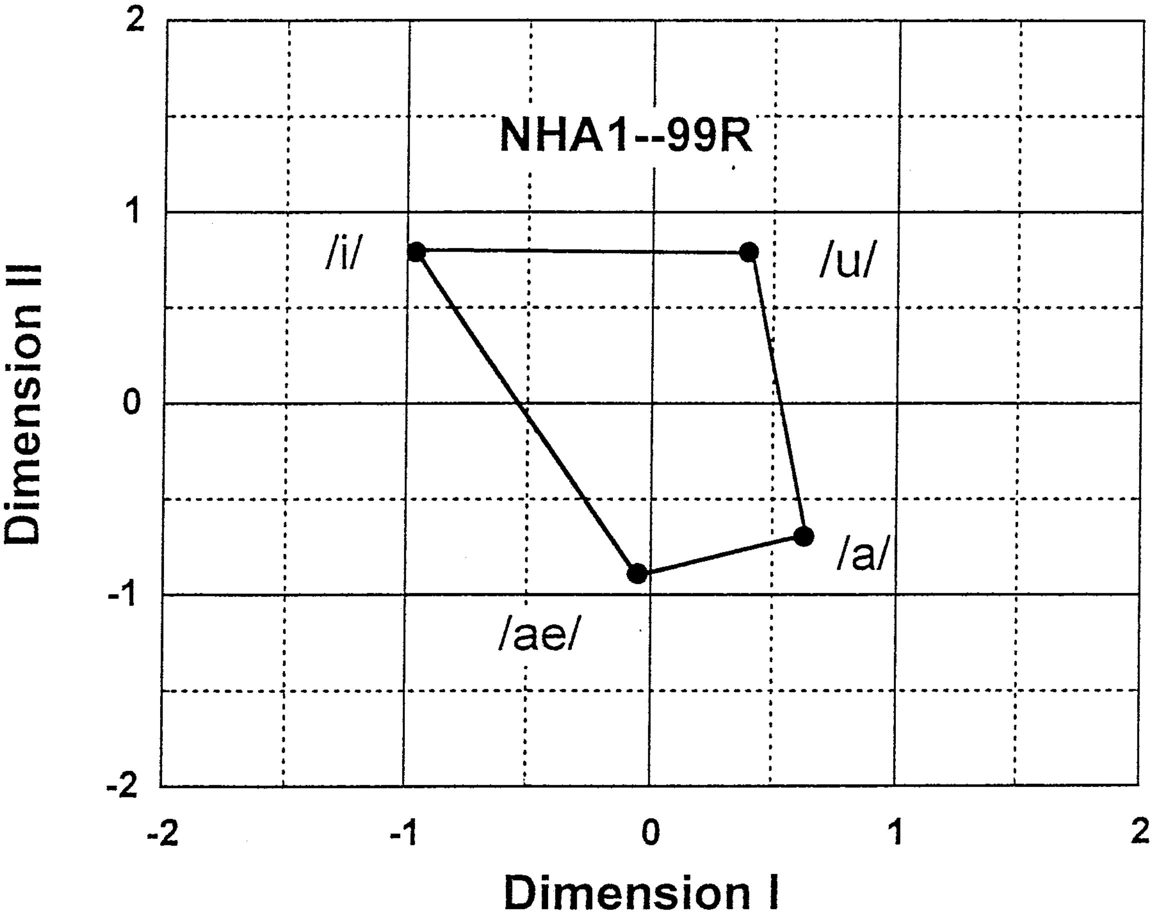

FIGURE 2

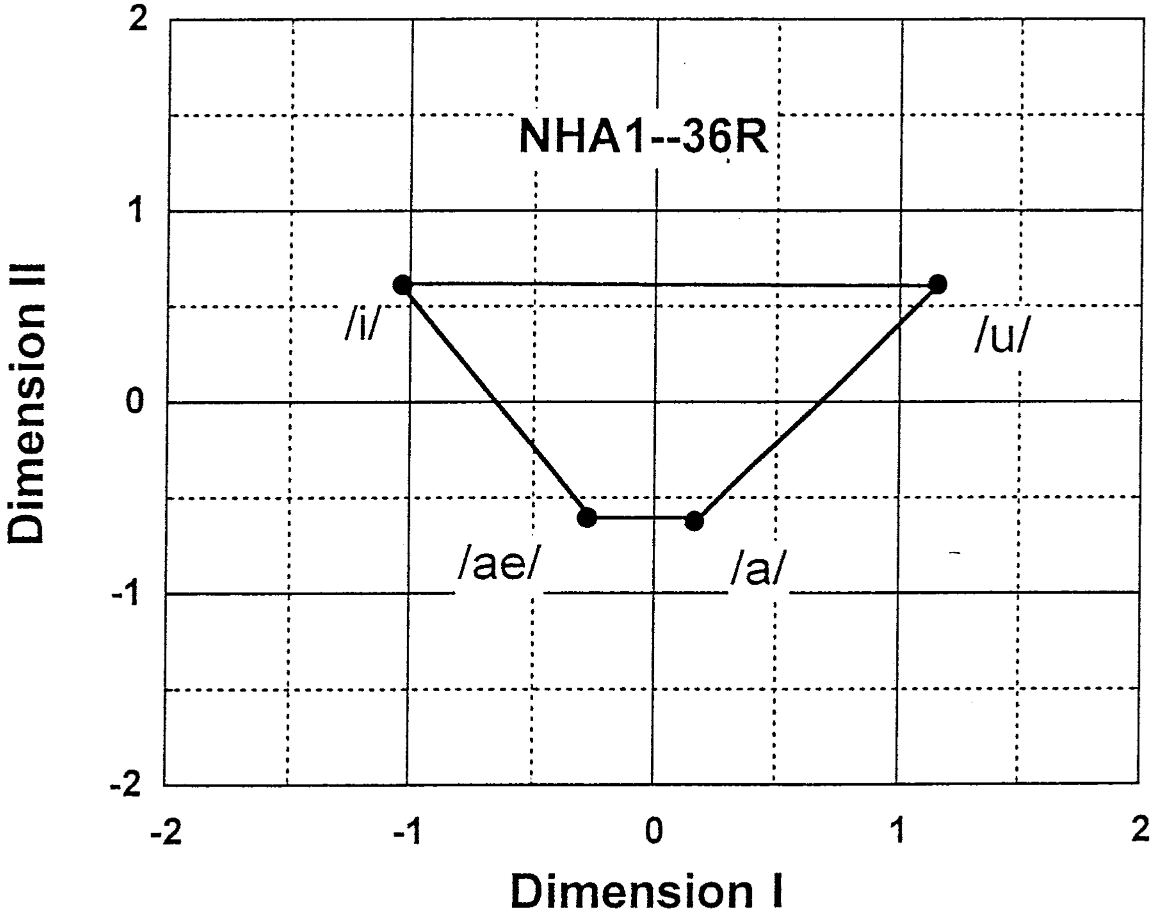

FIGURE 3

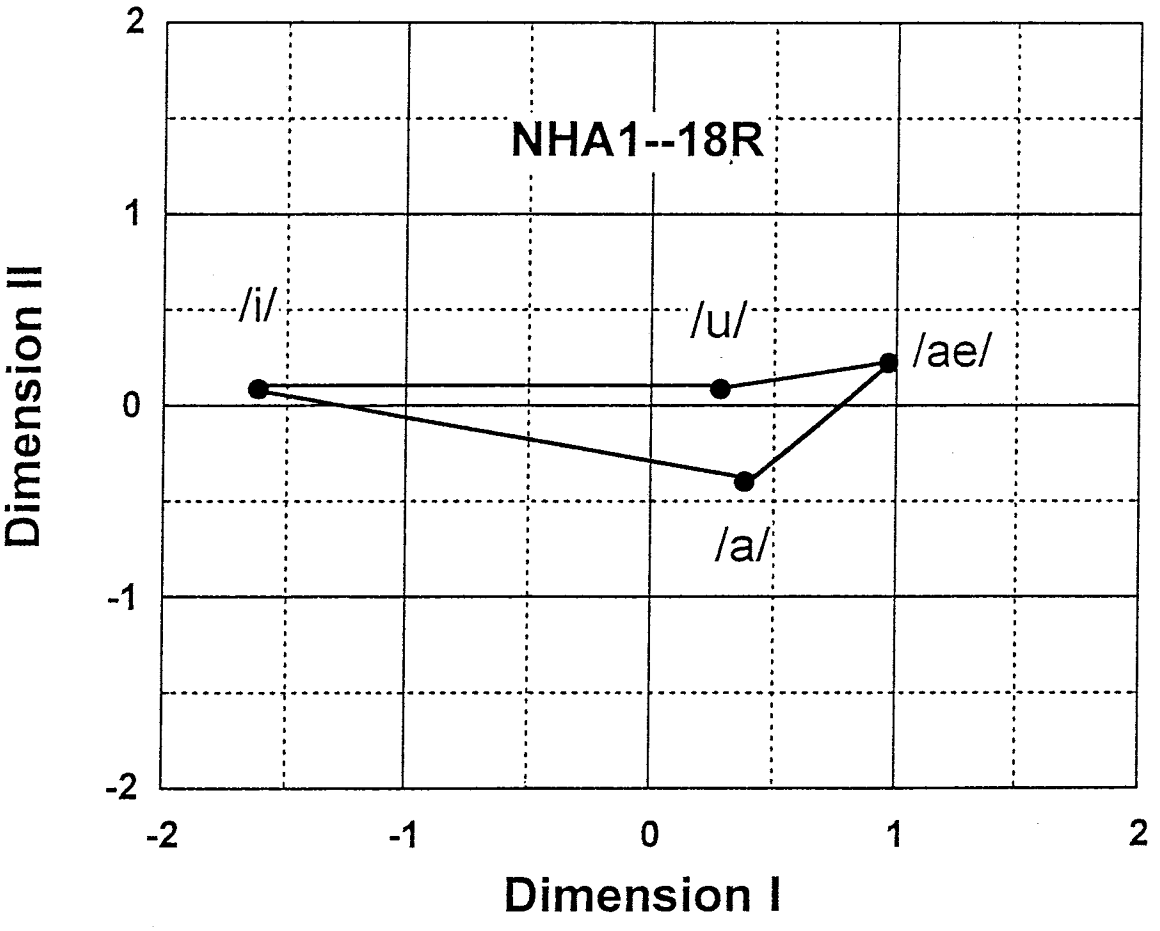

FIGURE 4

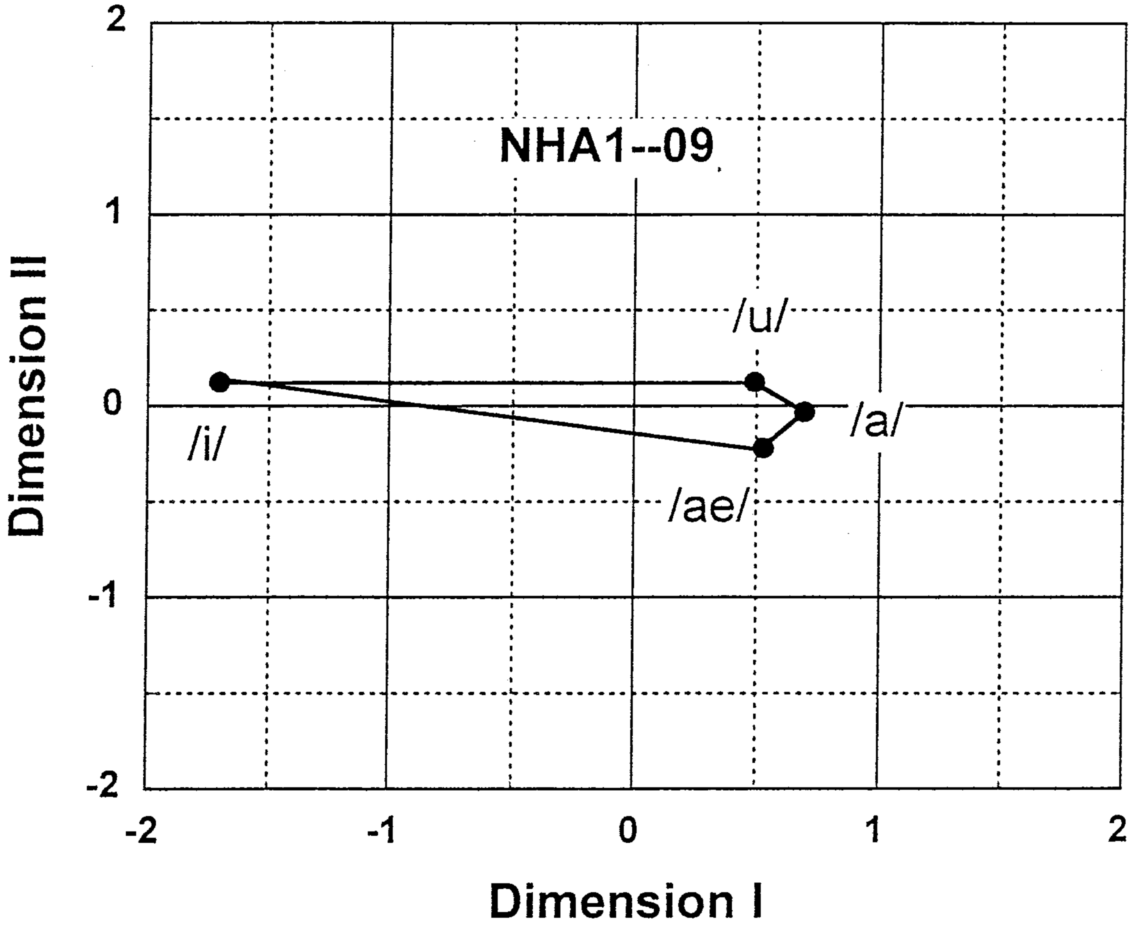

FIGURE 5

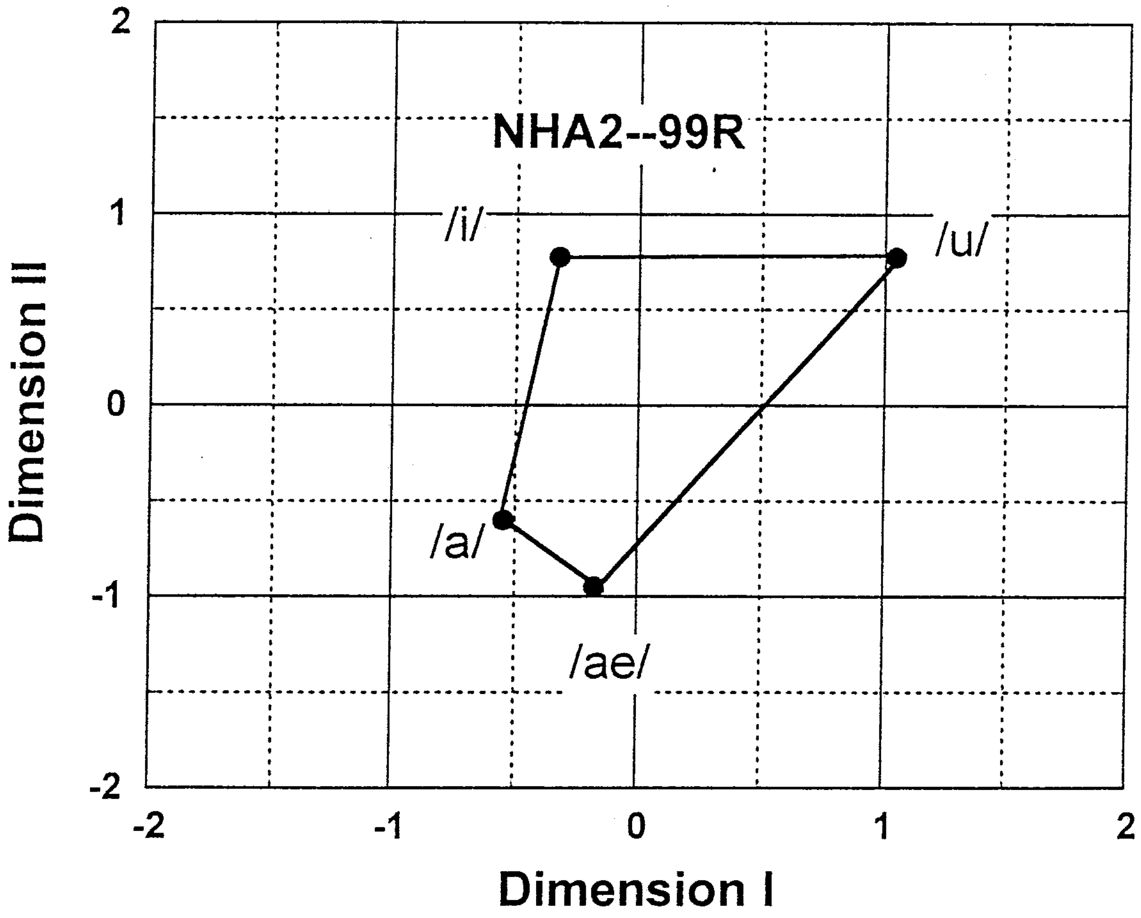

FIGURE 6

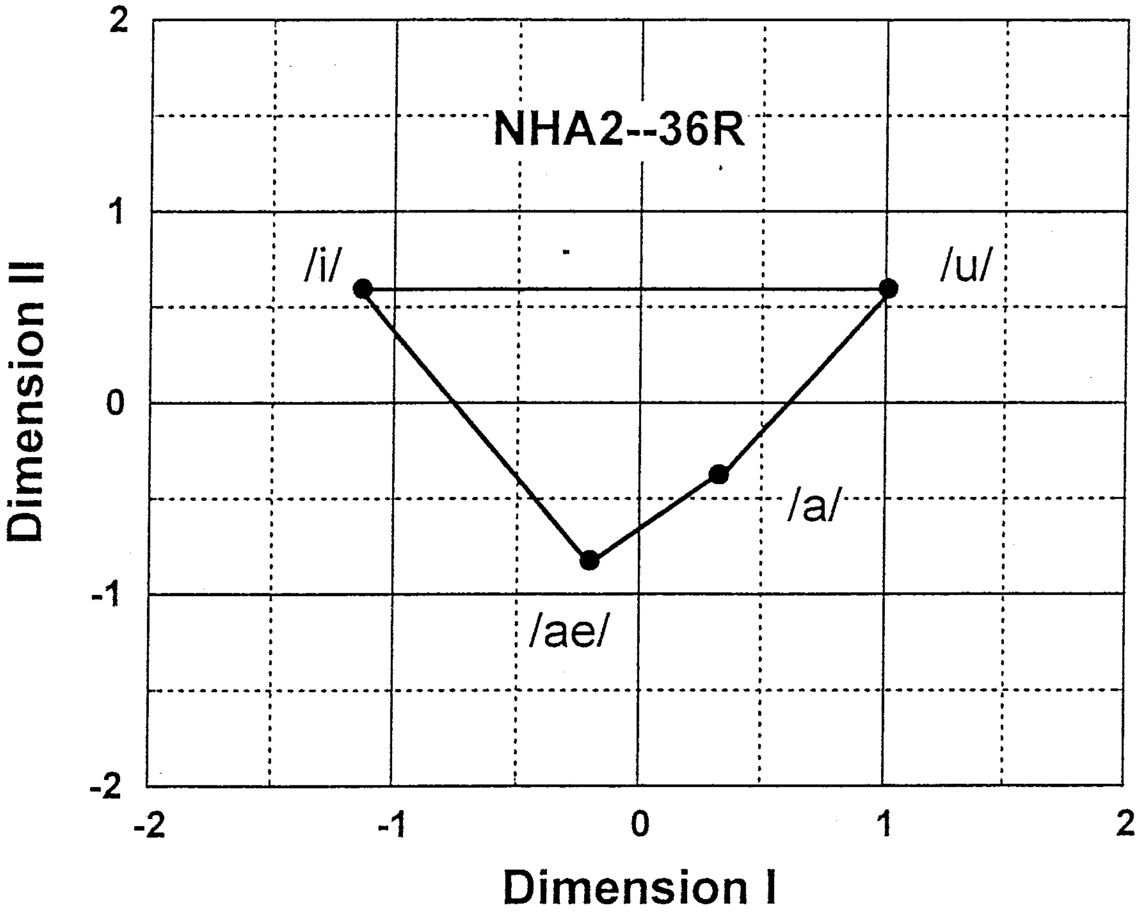

FIGURE 7

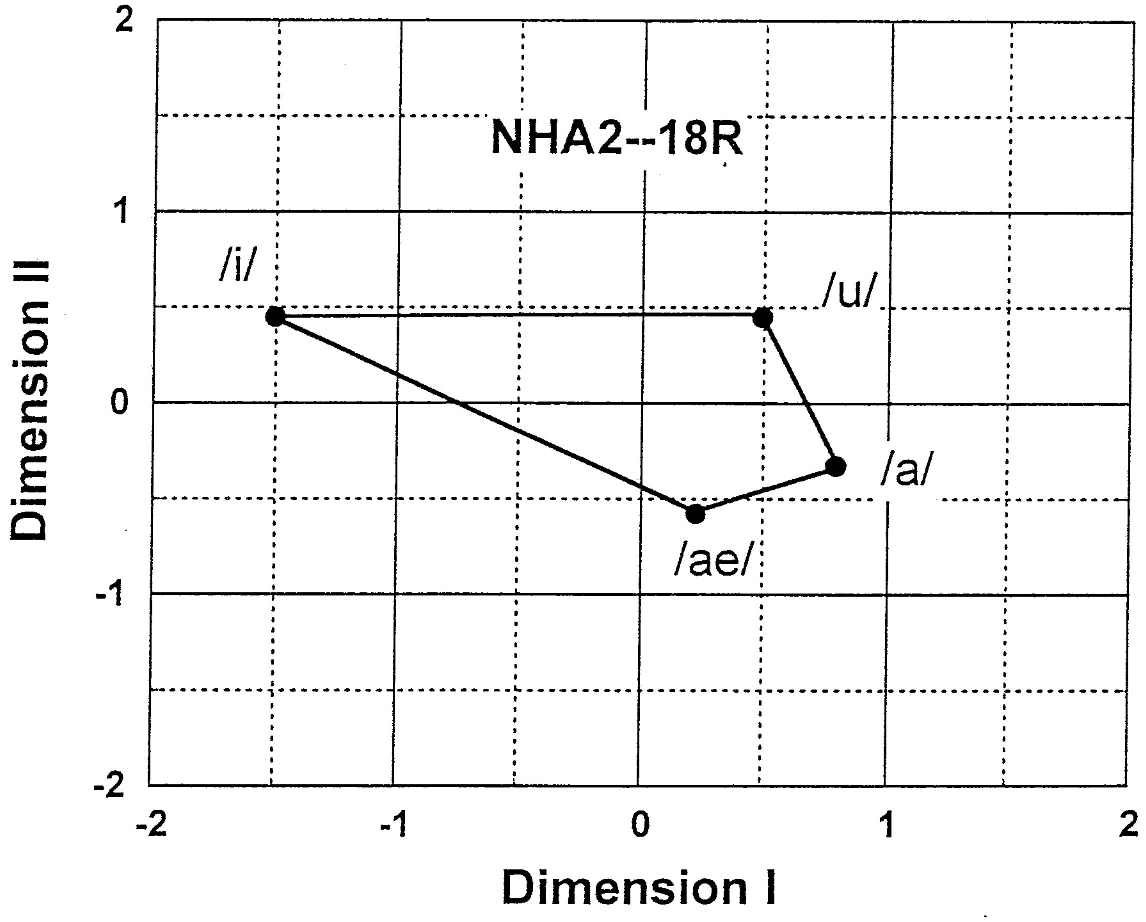

FIGURE 8

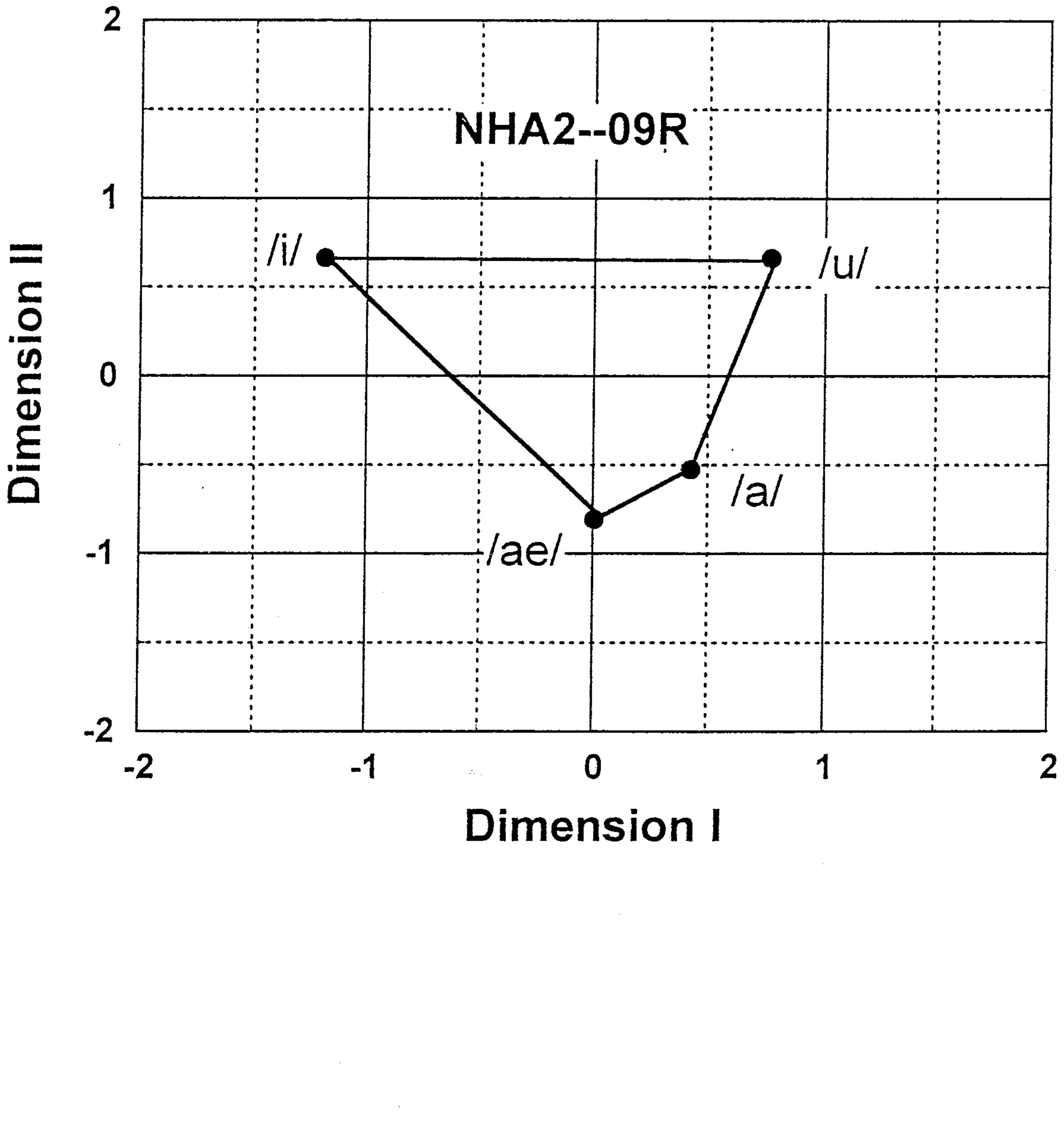

FIGURE 9

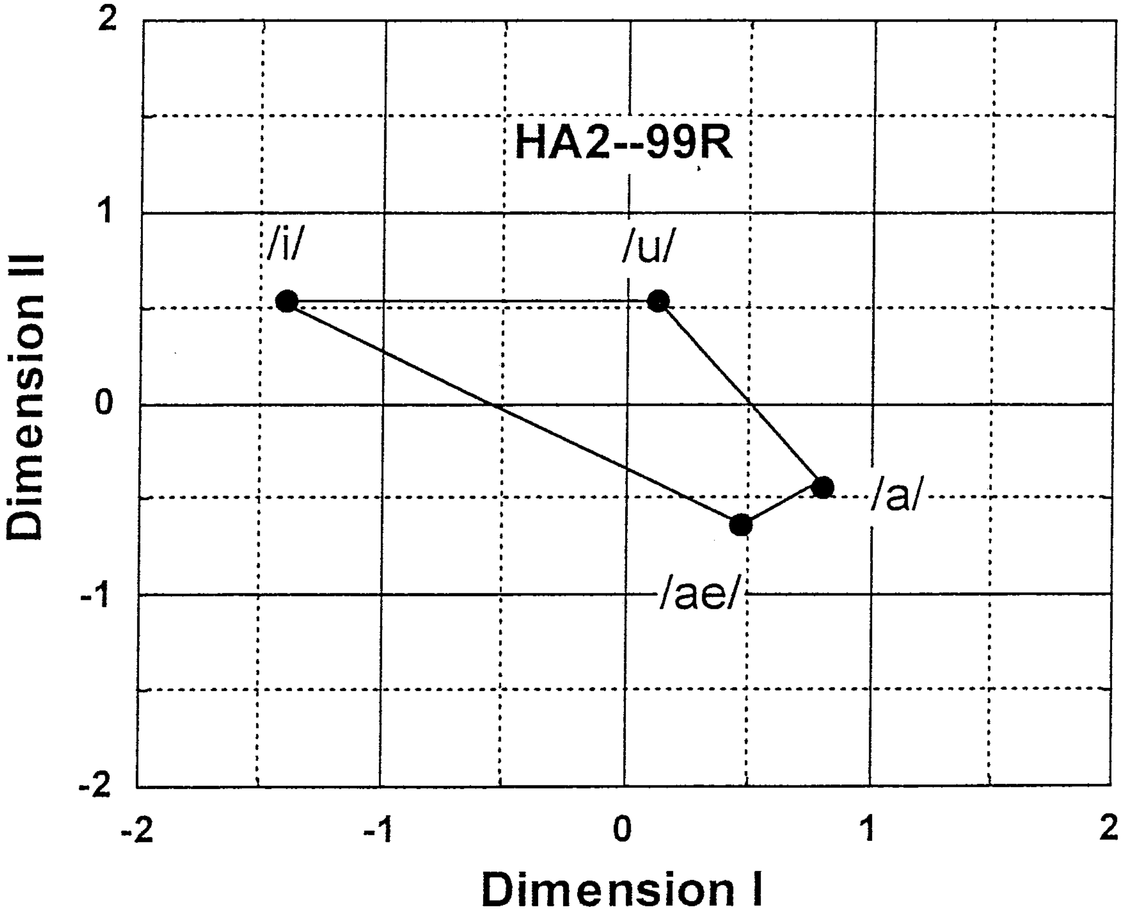

FIGURE 10

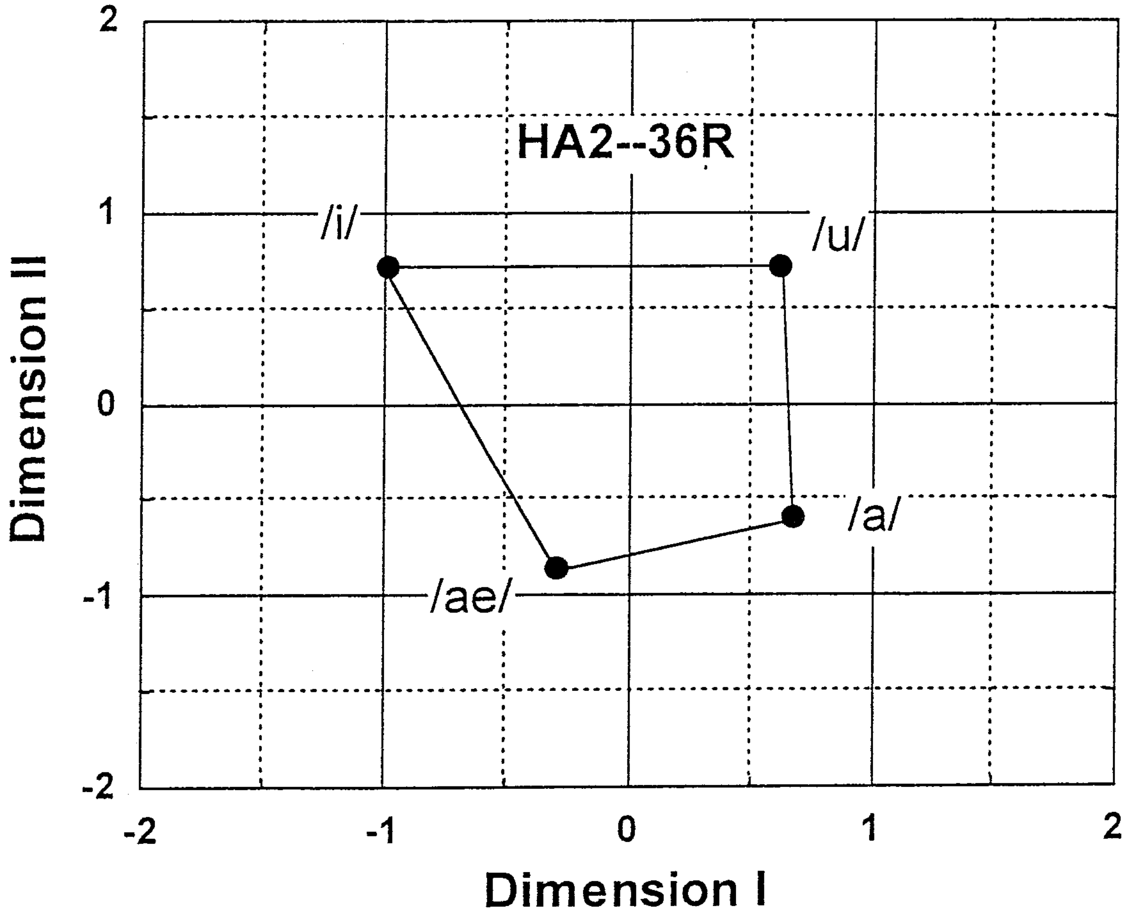

FIGURE 11

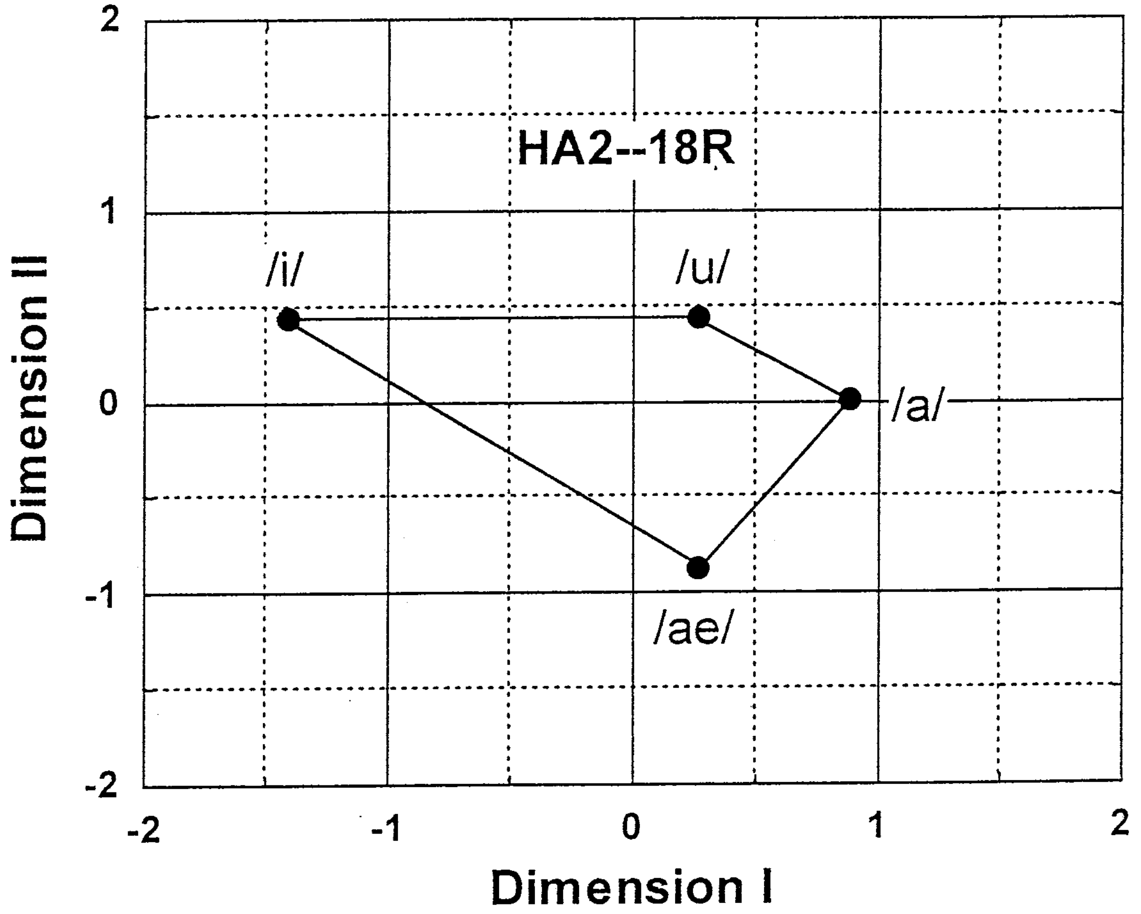

FIGURE 12

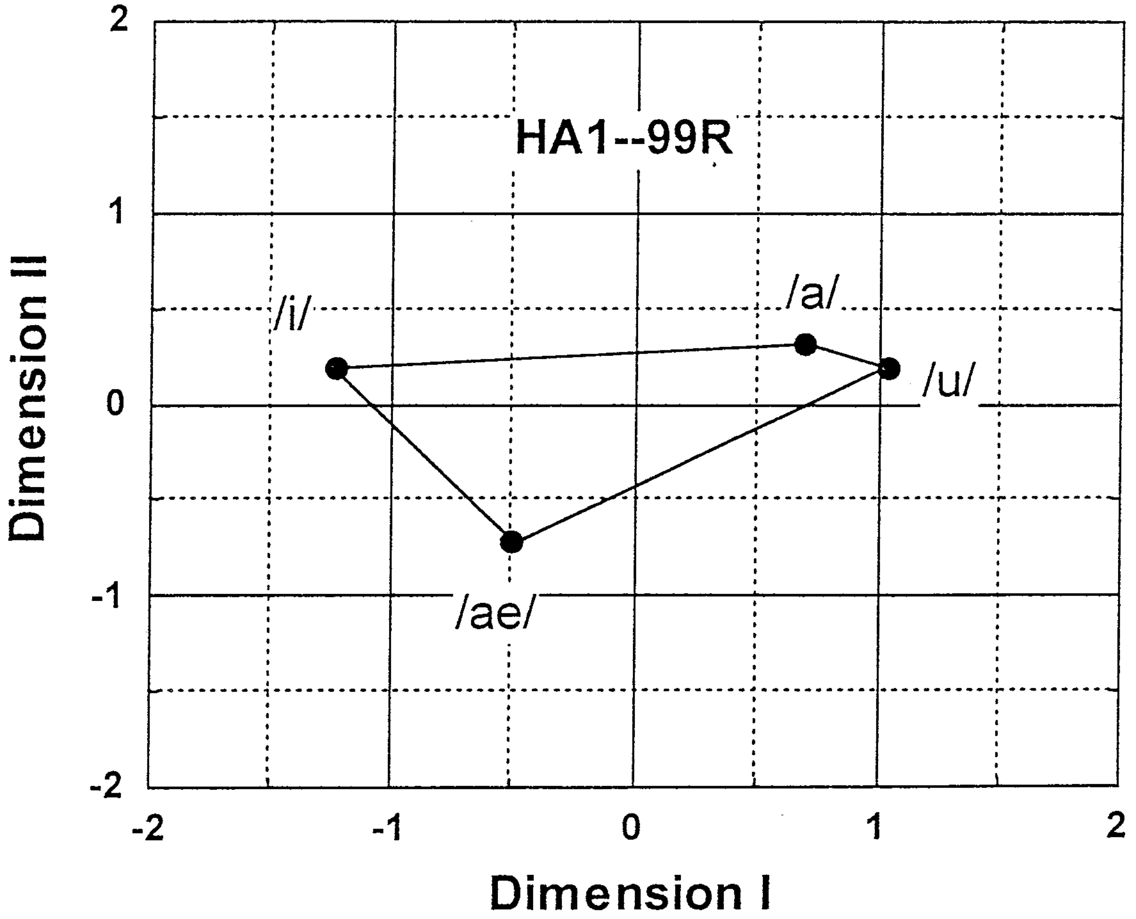

FIGURE 13

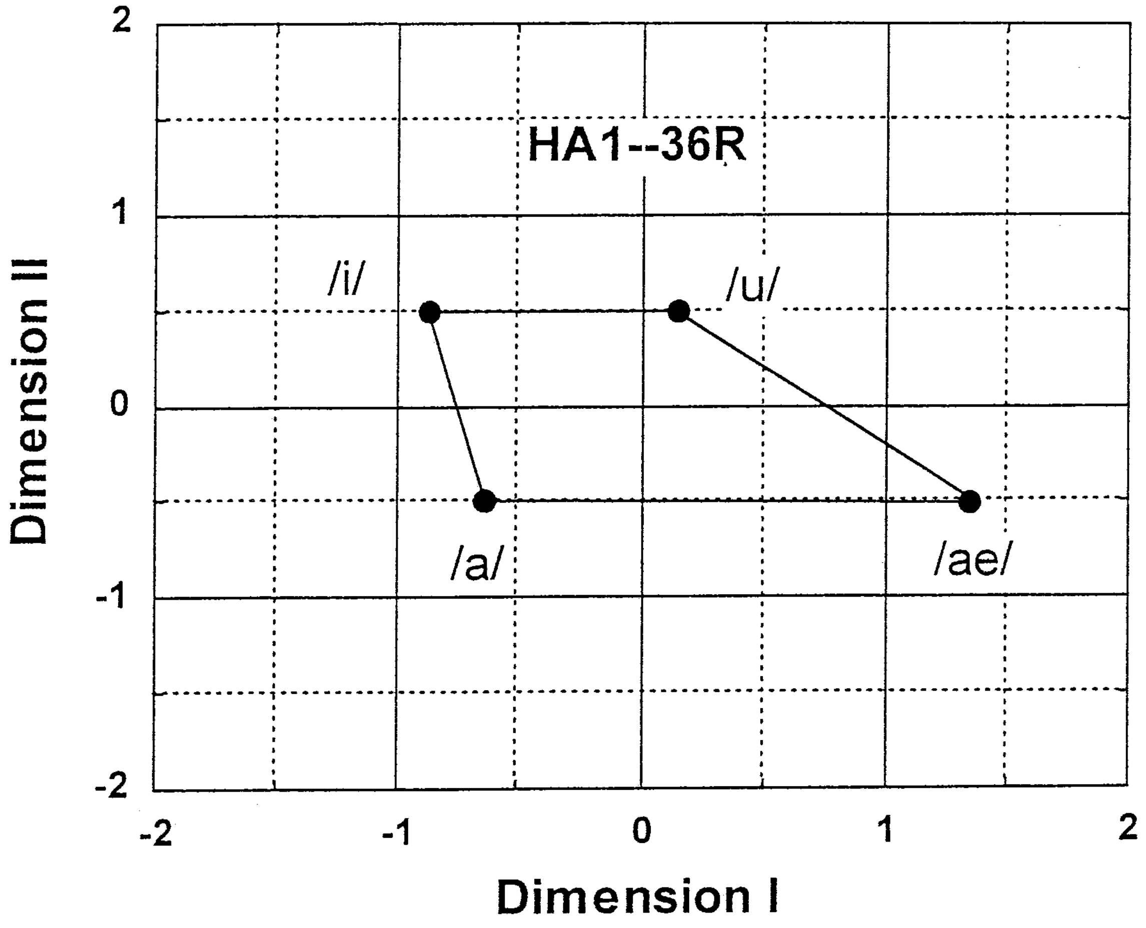

FIGURE 14

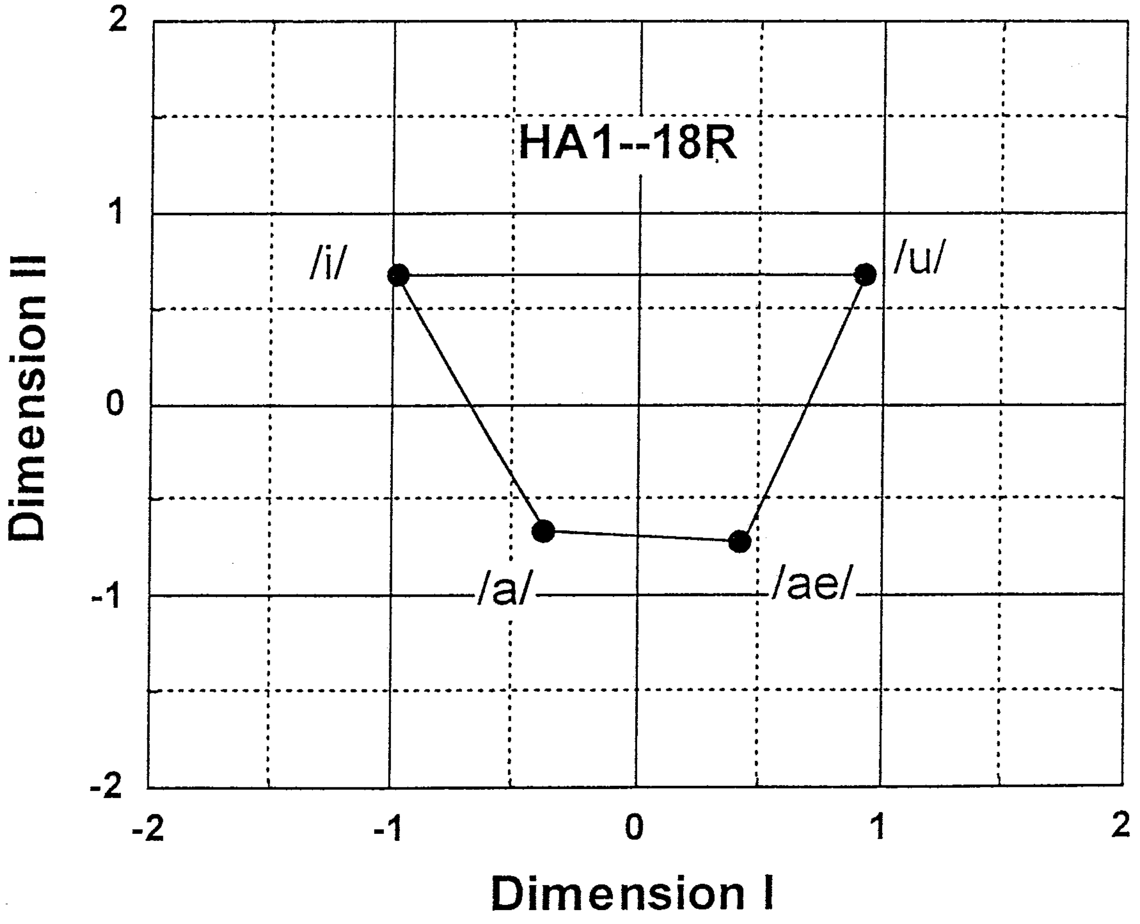

FIGURE 15

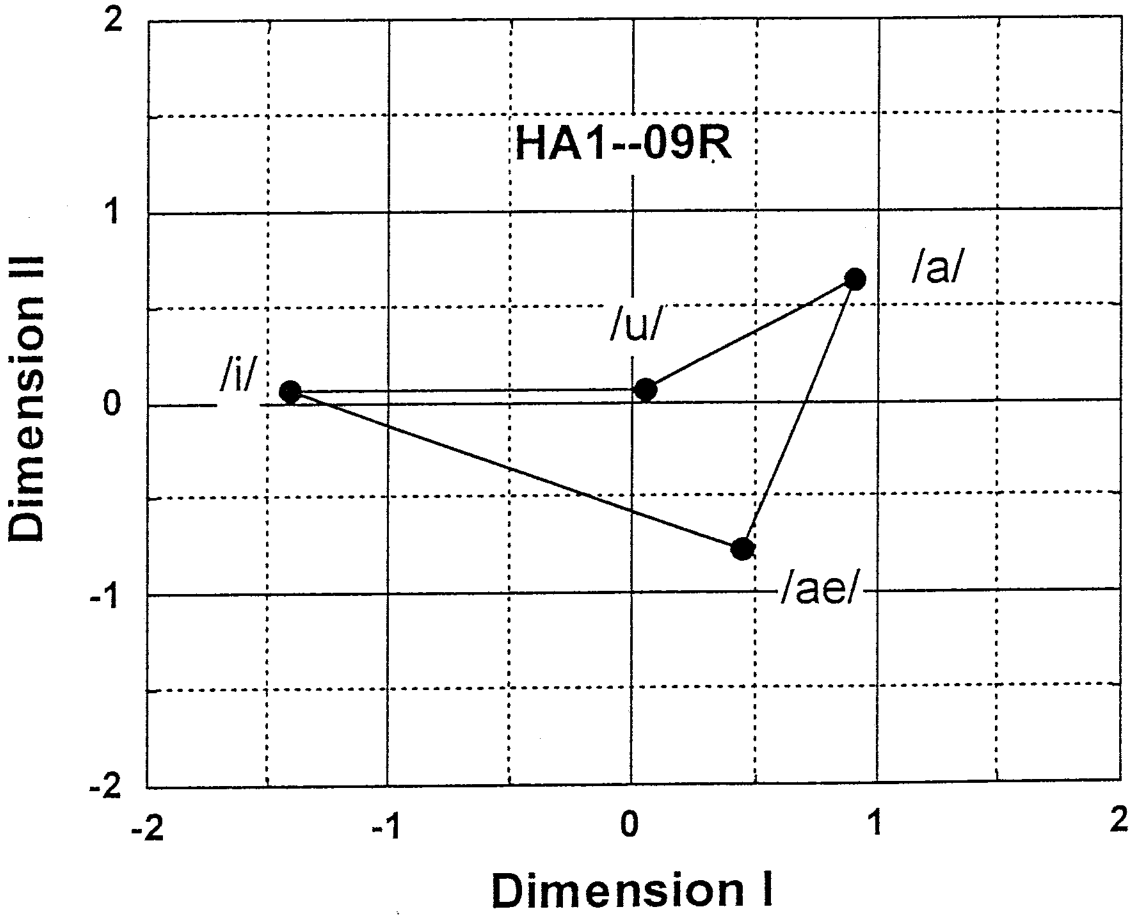

FIGURE 16

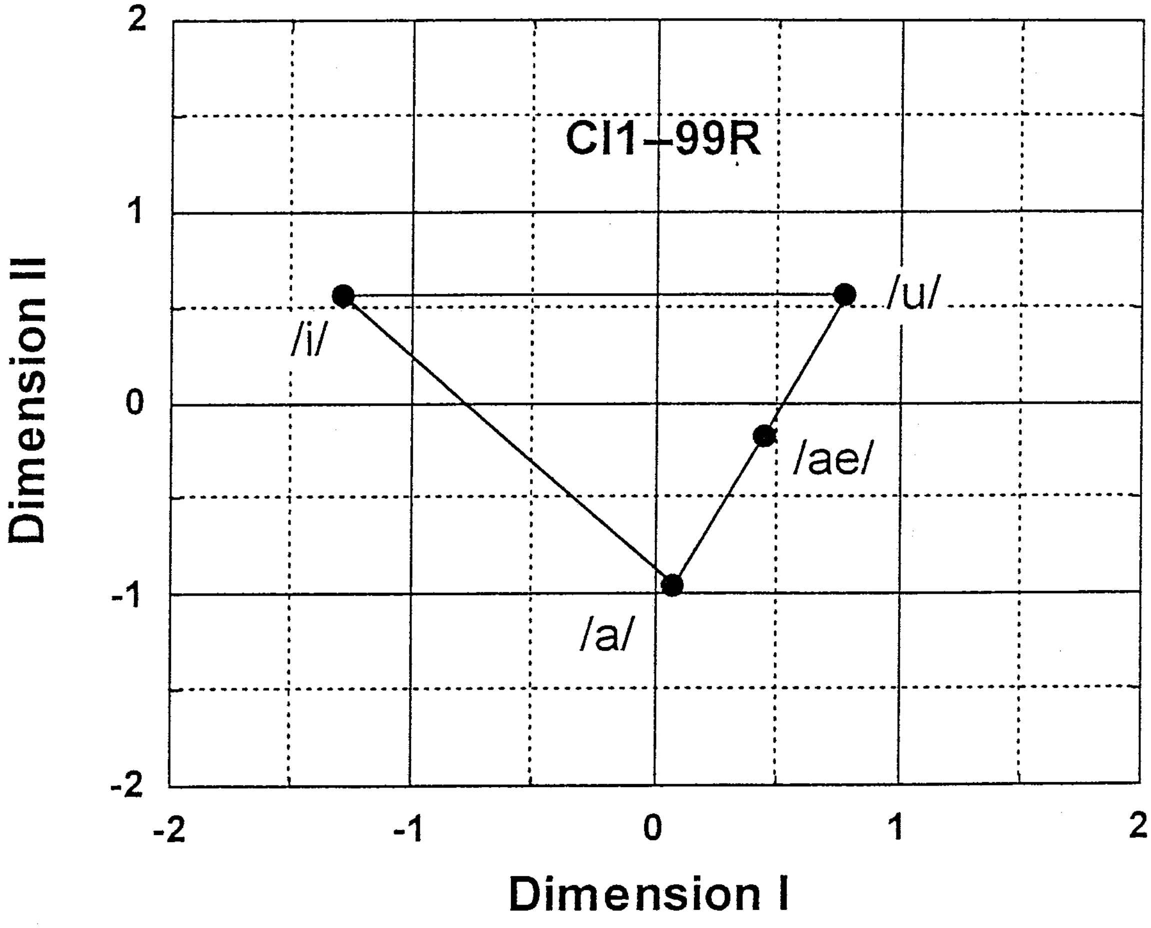

FIGURE 17

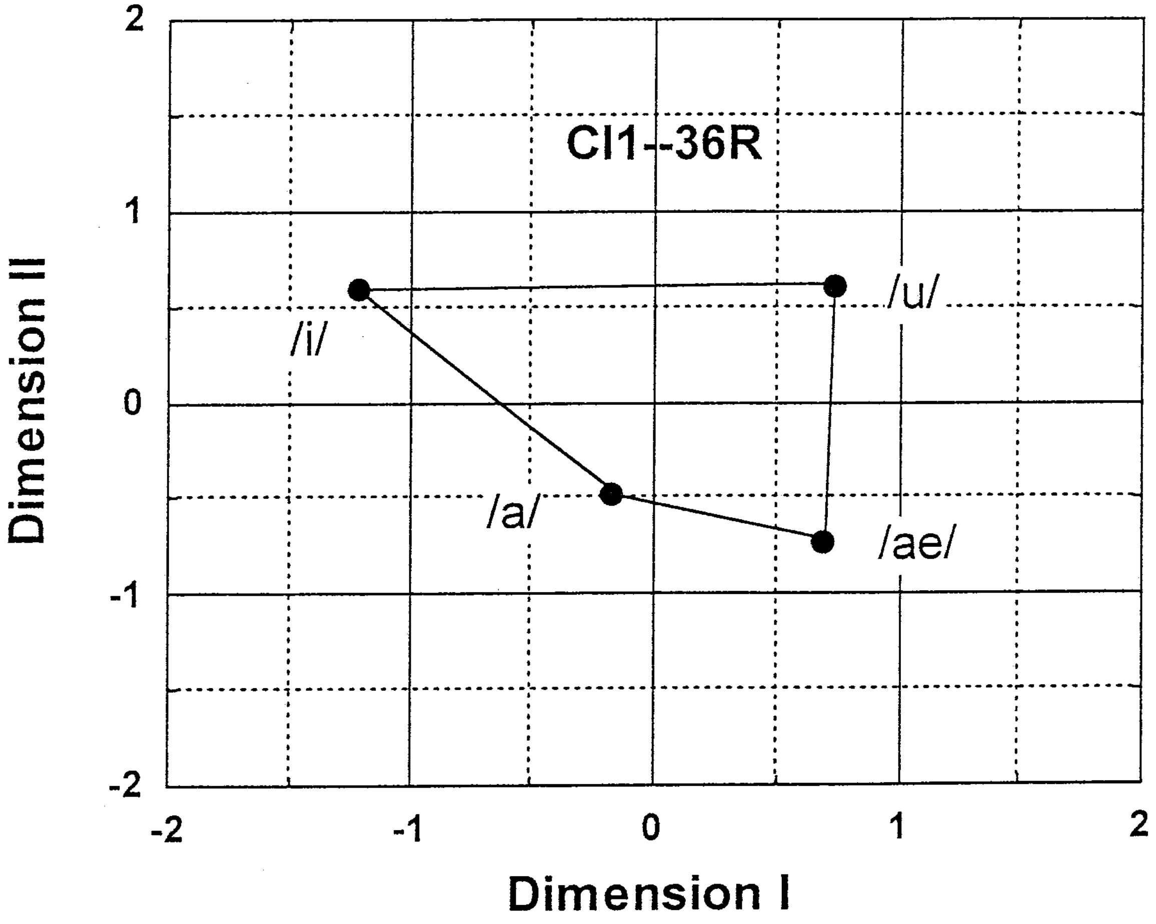

FIGURE 18

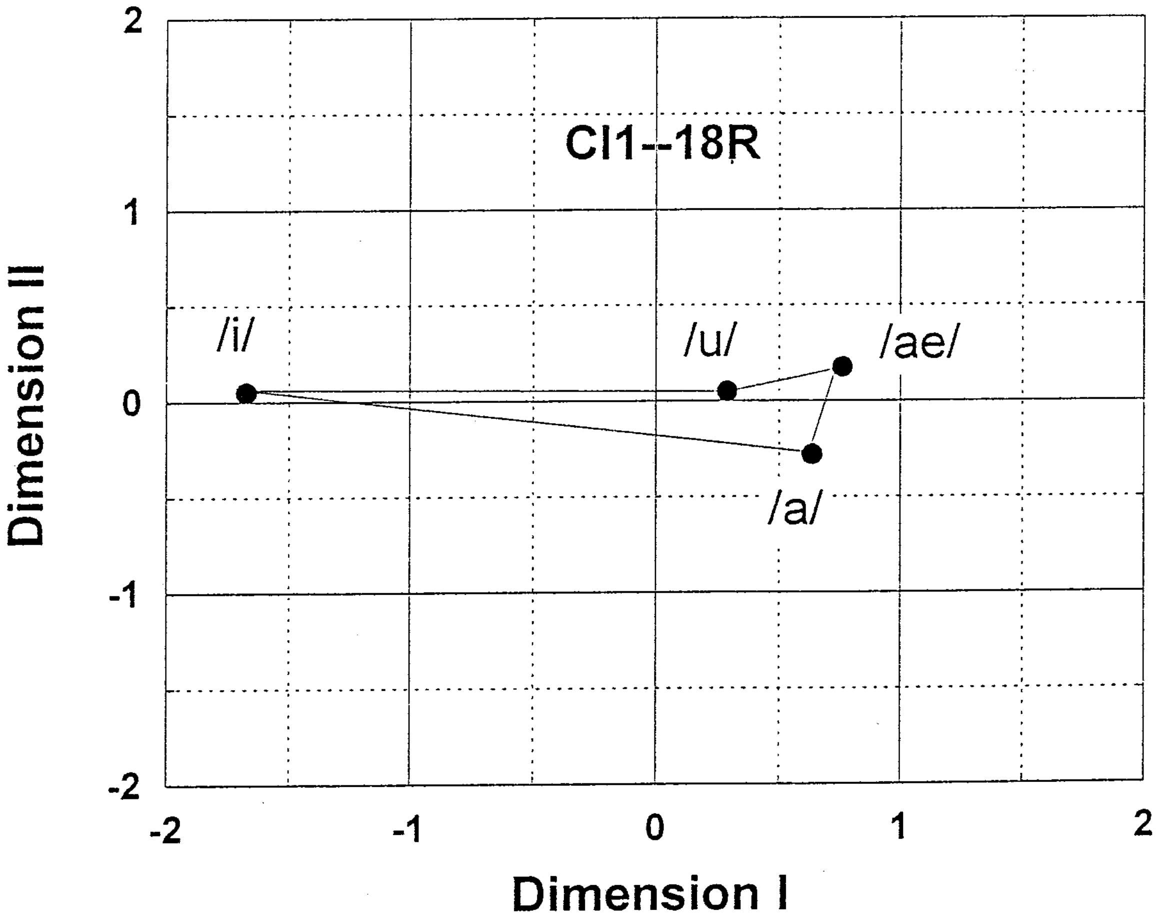

FIGURE 19

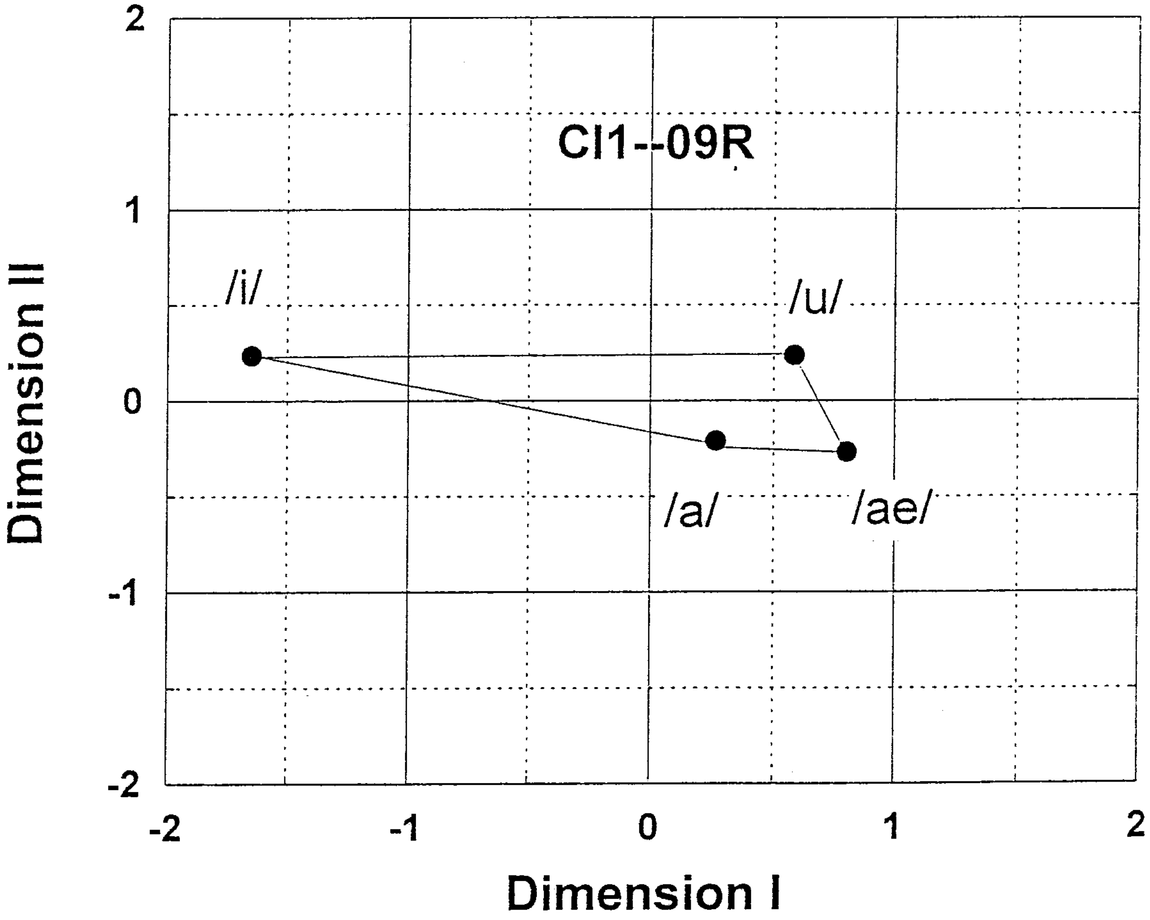

FIGURE 20

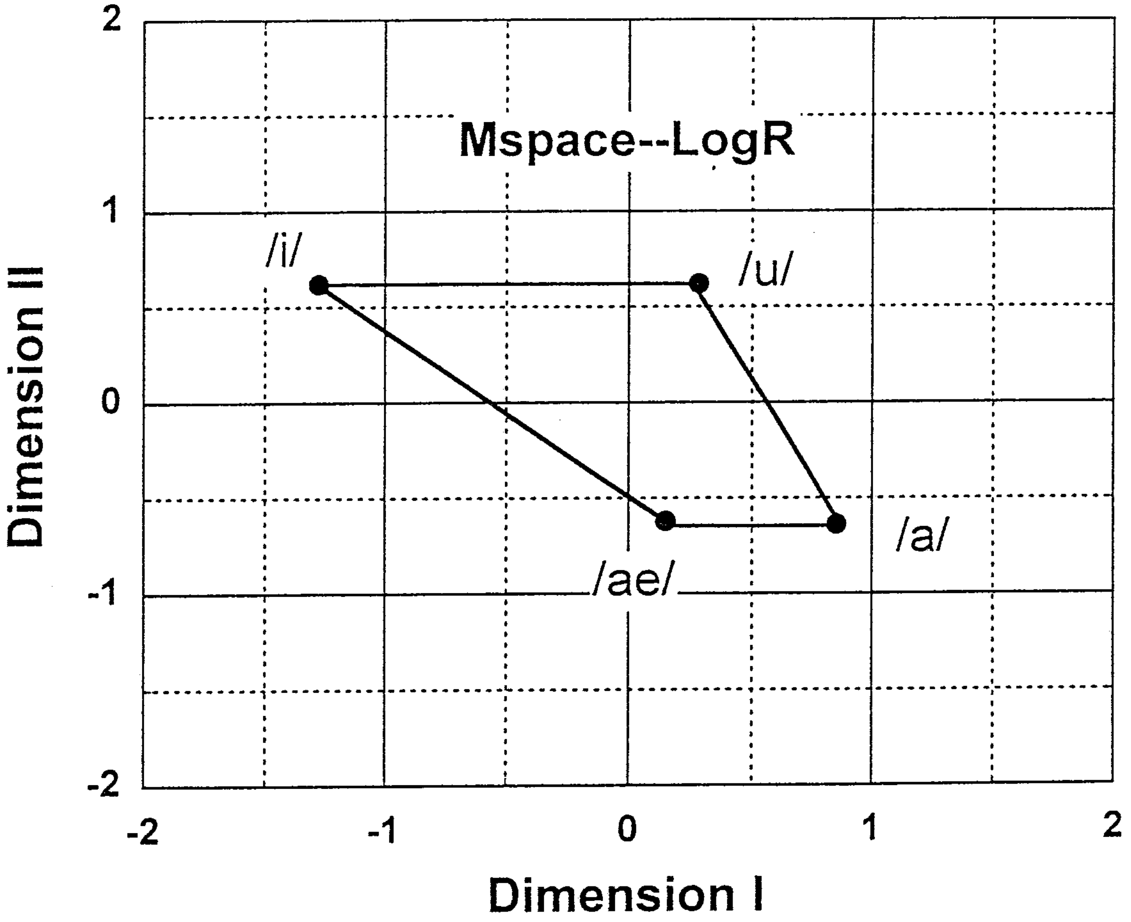

FIGURE 21

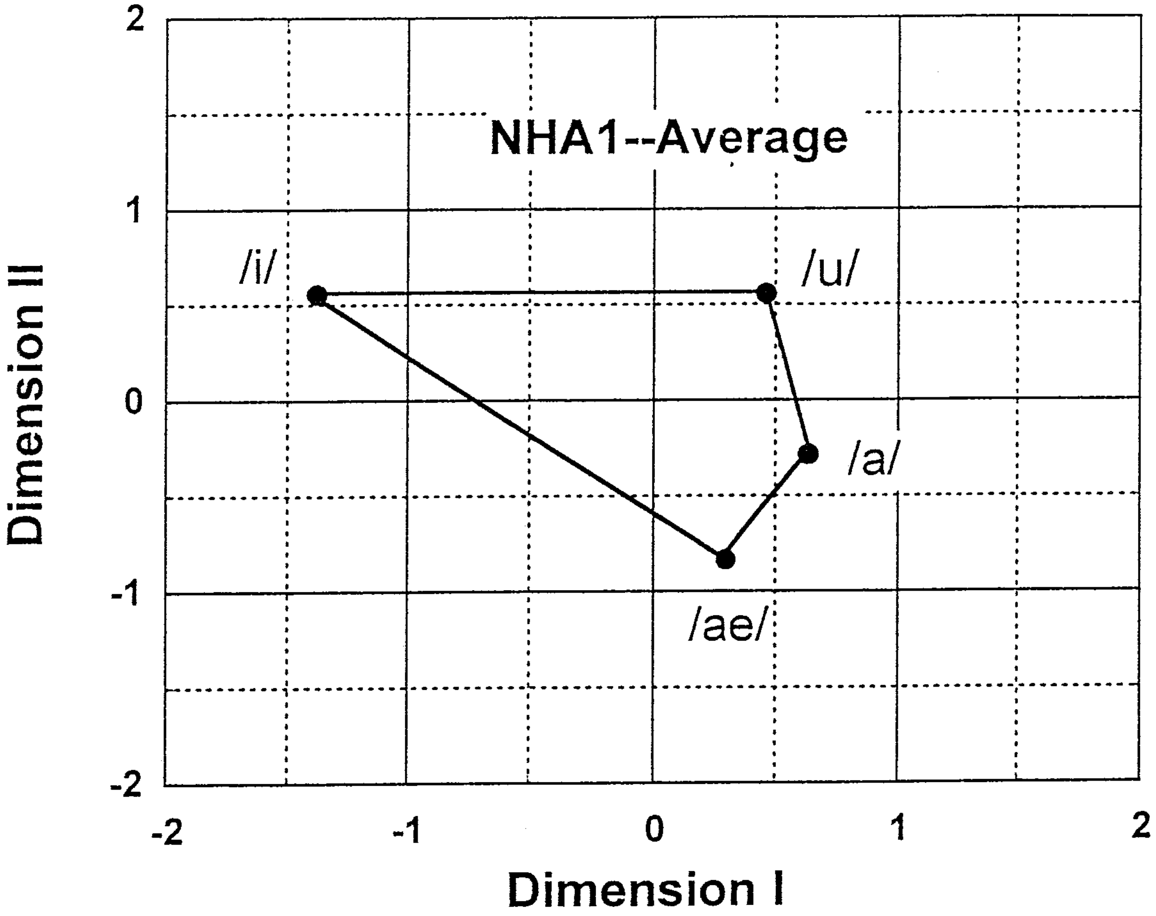

FIGURE 22

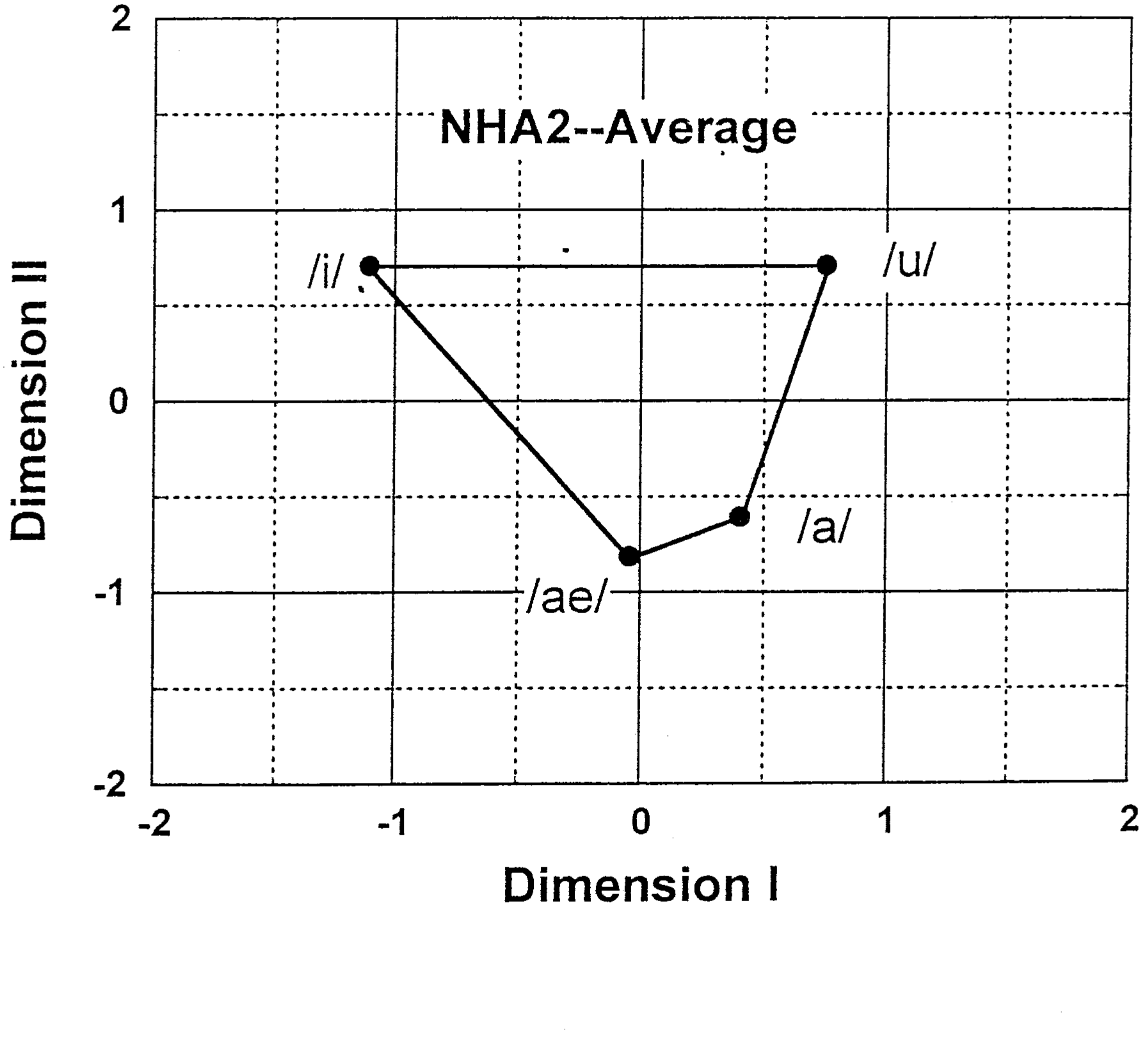

FIGURE 23

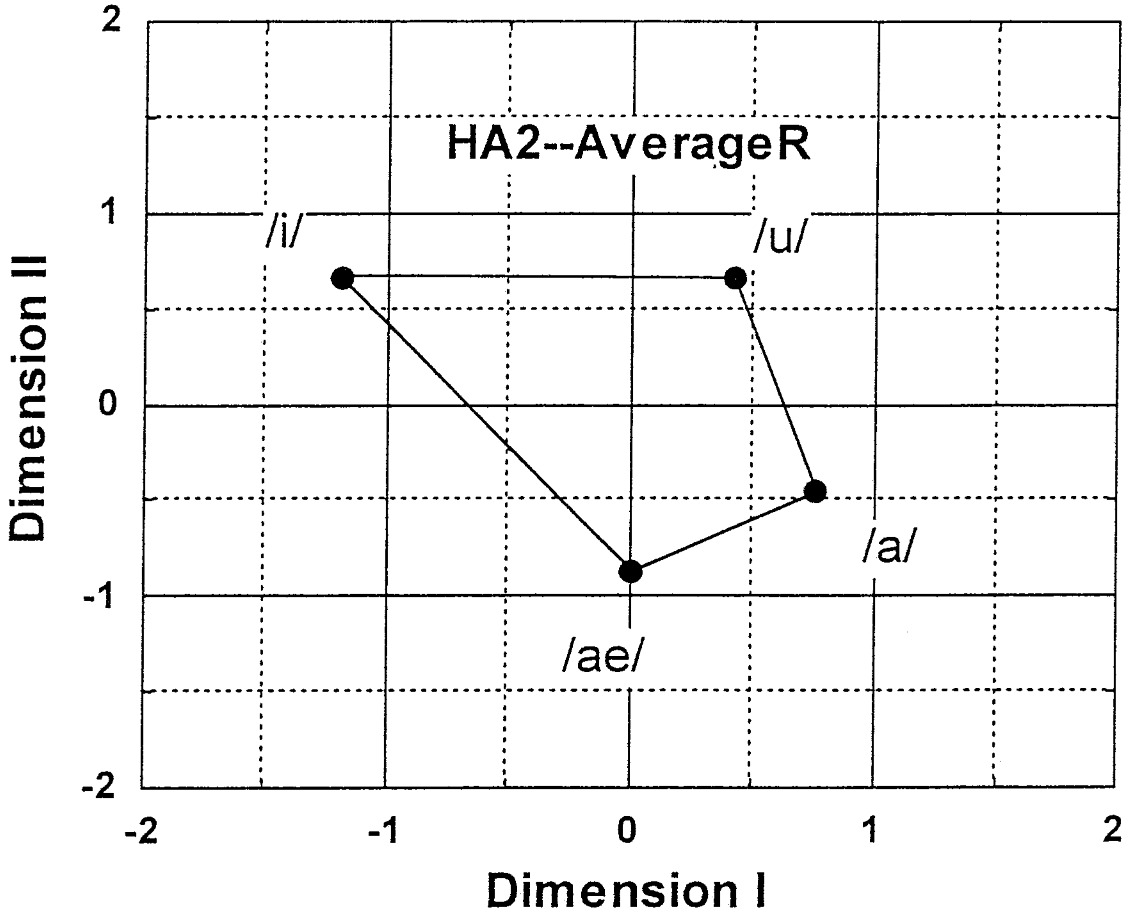

FIGURE 24

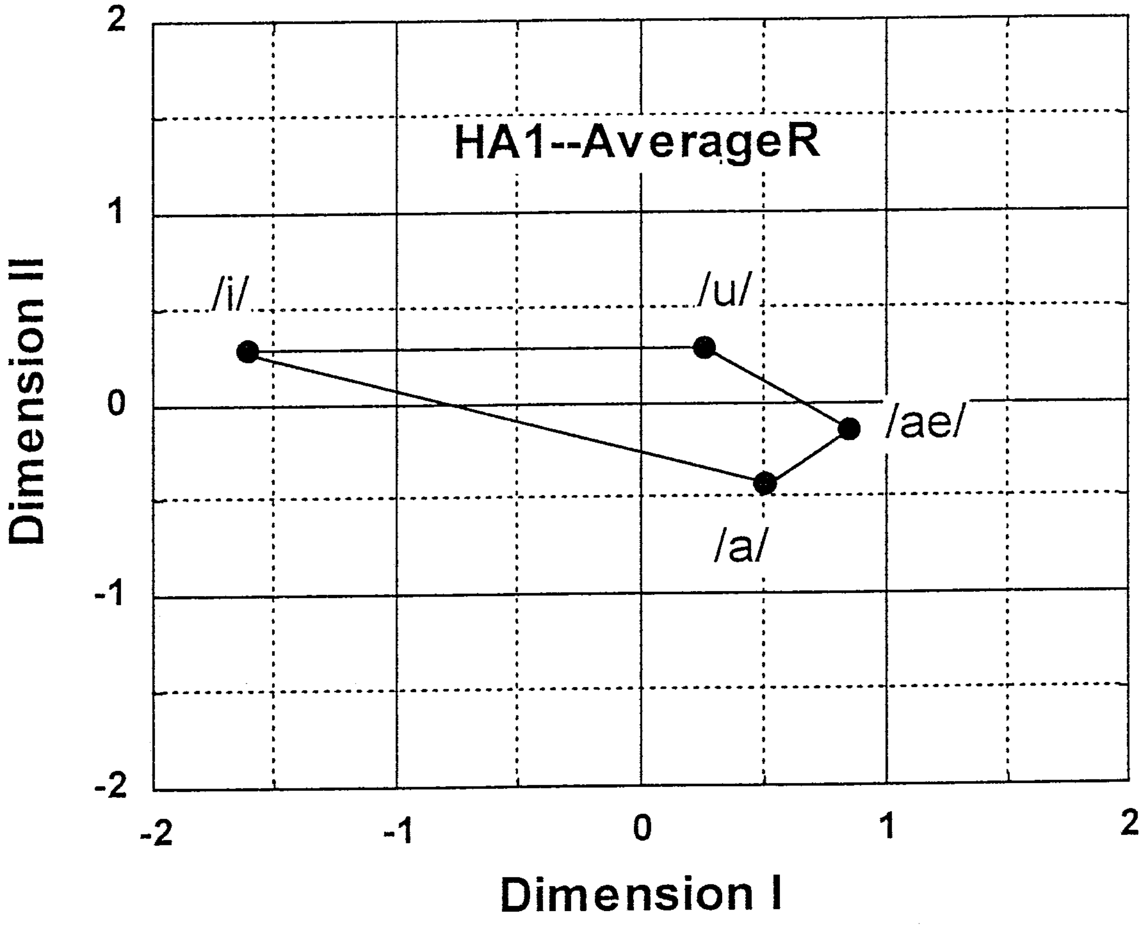

FIGURE 25