IMPLICATION OF THE PERCEPTION OF SIMILARITIES FOR PHONETIC THEORY

SADANAND SINGH AND DAVID R. WOODS

The word-expressions of a language are represented as sequences of phonemes. The phonemes of a language form a relatively small set of linguistic units which can be defined in terms of the invariant phonetic features which are realized in spite of great variation due to different contexts and speakers. Phonetic analysis has sought to define phonemes in terms of their articulatory or acoustic features. But the results of such analysis are not always unambiguous.

In recent years, a number of studies have been done which have tried to find the feature structure of phonemes without considering directly either their articulatory or their acoustic structure. These studies involved various perceptual tasks, such as identification of simple stimuli such as *pa*, *ta*, *ma*, etc. in noisy or distorted conditions, the recall of such stimuli in short-term memory, or direct judgments of similarity on pairs of such stimuli. These types of data have in common that errors in identification or in short-term memory recall of phonemic stimuli tend to correspond closely to the perceived similarities between pairs of phonemes and they can all be analyzed by such techniques as multi-dimensional scaling or factor analysis. The resulting perceptual dimensions or factors can usually be related to articulatory or acoustic features and serve to support or refute particular phonetic analyses of the phoneme of a language.

In the present study, some results are reported on the perception of English vowels and consonants by native speakers of English as well as on the perception of English, Hindi, and Korean stops by native speakers of each of these three languages. Using multi-dimensional scaling techniques, these studies have analyzed direct judgments of similarity on pairs or triples of simple phonemic stimuli, e.g., consonants in initial position before the vowel /a/ or vowels in the context of initial /h/ and final /d/.

Two studies on American English vowel phonemes found that the structure of vowel perception could be closely related to traditional articulatory dimensions of tongue height and tongue advancement or to the corresponding acoustic dimensions of first and second formants respectively (Singh and Woods 1971, Anglin 1971). Figure (1) plots 12 English vowels which were heard in the context of initial /h/ and final /d/. The front vowels are on the right, the back vowels on the left; the high

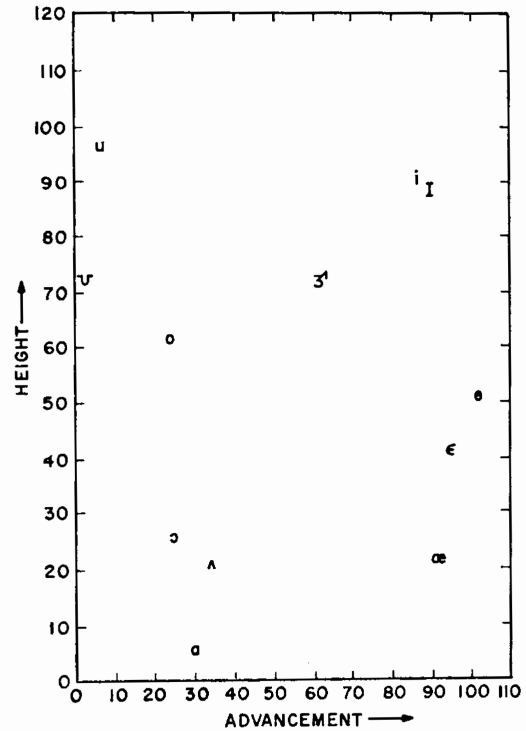

Fig. 1. Advancement/Height.

vowels are above and the low vowels below the figure. Two further dimensions were recovered from the vowel data. Figure (2) plots these dimensions. The vowel /ɝ/ in the American English pronunciation of HEARD appears to be opposed to the remaining eleven vowels on the dimension interpreted as 'retroflex'. The fourth dimension was interpreted as 'tenseness'. A similar study done on the same vowels but spoken in isolation failed to recover a tenseness dimension (Singh and Woods 1971). This result can be explained by the fact that the non-tense vowels cannot occur in isolation in English and hence may not be realizations of English phonemes.

Three studies were conducted on 22 American English consonants. Using Carroll and Chang's INDSCAL technique, five dimensions were recovered which were interpreted as sibilance, place of articulation, plosiveness, voicing, and nasality. The first dimension, sibilance, distinguishes the consonants /s z ʃ tʃ dʒ/ from the remaining; its interpretation as sibilance rather than stridency is due to the fact that /f/ and /v/ do not join this cluster. The second dimension, place of articulation, only effectively divides front from back consonants. No consistent division of the front consonants into labials and alveolars was recovered from these data. The third dimension, plosiveness, reflects manner of articulation, but, unlike other analyses of perceptual data (Singh, Woods and Tishman 1972) does not distinguish the non-plosive obstruents from the sonorants.

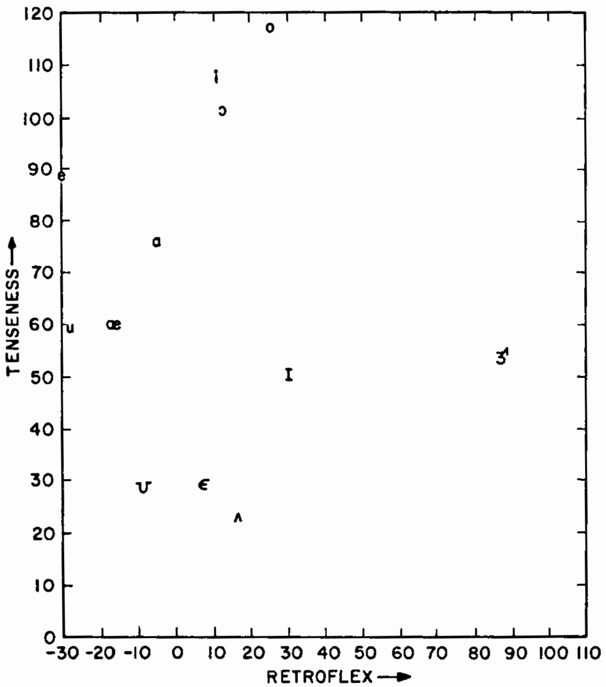

Fig. 2. Retroflex/Tenseness.

The cross-language studies of the English, Hindi, and Korean stops supported the hypothesis that listeners' perceptions are in part a product of their linguistic backgrounds. For the six English stops the three places of articulation and voicing were recovered from the data. The nine Korean stops showed the same places of articulation as English plus a manner dimension which distinguished aspirates, plain, and glottalized stops from each other. The 16 Hindi stops, again, showed three places of articulation but had additional dimensions of retroflexion, aspiration, and voicing. The three listening groups weighted the dimensions of the English stops similarly, perhaps because their phonetic features corresponded to features of the Korean and Hindi stops. But the unique retroflexion dimension of Hindi stops and glottalization dimension of Korean stops received greater weights when perceived by the native listeners of these languages.

In general, the multi-dimensional analysis of similarity judgments leads to the recovery of perceptual dimensions of consonants and vowels that can be related to articulatory or acoustic analysis. In a few cases, however, the results may support one phonetic analysis over another. For example, the sibilants /s z ʃ tʃ dʒ/ cluster together but stridents /f/ and /v/ do not join this group, thus favoring the choice of the feature sibilant over strident.

Place of articulation remains an interesting problem. The analysis of 22 English consonants results in the recovery of only a front/back dimension whereas the small

sets of English, Hindi, and Korean stops permit the recovery of three places of articulation. This may be due to the fact that the stops fall into subsets with identical places of articulation, e.g., bilabial, apico-alveolar, etc., whereas the full set of 22 consonants are less well defined, e.g., the labials may be bilabial or labio-dental, the apicals may be interdental or alveolar, etc.

The multi-dimensional analyses also raise some questions about the relations between features, not brought up in phonetic analyses. In particular some feature dimensions are stronger than others, e.g., the back/front dimension in both vowels and consonants is the strongest. Furthermore, certain feature specifications seem to increase perceived similarity of pairs of phonemes when the specification is common to both members of the pairs, e.g., a nasal pair /m-n/ is perceptually more similar than the parallel oral /b-d/.

Speech Department
Howard University
Washington, D.C.

REFERENCES

Anglin, M.N.
1971 "Perceptual Space of English Vowels in Word-Context", Masters thesis, (Howard University).

Singh, S., and D.R. Woods
1971 "Perceptual Structure of 12 American English Vowels", *Journal of the Acoustical Society of America* 49.6 (Part 2):1861-1866.

Singh, S. D.R. Woods, and A. Tishman
1972 "An Alternative MD-SCAL Analysis of the Graham and House Data", *Journal of the Acoustical Society of America* 51.2 (part 2):666-668.

DISCUSSION

BLACK (Columbus, Ohio)
I would appreciate a further explanation of the coordinates of the vowel-graphs, particularly of the abscisses. What units can these be stated in? Are they linear?

SINGH
The coordinates are on an arbitrary scale, but the scale is the same for each dimension in the perceptual space. The important thing is the relative interpoint distances.

LADEFOGED (Los Angeles)
Could you give more details of the analysis programs you used? For example, were the factor axes orthogonal?

SINGH

The vowels were analyzed using MD-SCAL program of Sheperd and Kruskal and a rotation program of Schönemann and Carroll. The consonants were analyzed by the IND-SCAL program of Carroll and Chang. The axes in each case were orthogonal.

FISCHER-JØRGENSEN (Copenhagen)

What were the types of questions to which the listeners should react?

SINGH

The methods included: (1) equal appearing interval scale, (2) magnitude estimation, and (3) ABX. In the first one, seven points were used in which one represented greatest similarity and seven greatest dissimilarity. In the second one, the subjects were asked to draw a horizontal line to estimate the similarity of a pair of phonemes: a shorter line for a more similar pair and a longer line for a more dissimilar pair. No standard was provided by the experimenter. In the third experiment three sounds were heard in sequence, e.g., /pa/, /ba/, and /sa/. The subjects were asked to judge whether the first one was more similar to the second or to the third ones.