PERCEPTUAL DIMENSIONS OF CANTONESE TONES: A MULTIDIMENSIONAL SCALING REANALYSIS OF FOK'S TONE CONFUSION DATA

JACK GANDOUR

1. INTRODUCTION

It is a common assumption in research on speech perception that a listener's internal representation of the speech signal is organised, at least in part, in terms of phonetic and/or phonological features employed in linguistic analysis. This process of speech perception has been summarised by Studdert-Kennedy (1975:253) as follows:

> In short, perception entails the analysis of the acoustic syllable, by means of its acoustic features, into the abstract perceptual structure of features and phonemes that charcterize the morpheme.

Many investigators have attempted to determine the number and nature of these perceptual dimensions or features that listeners put together in the identification of speech sounds - consonants (Singh 1975), vowels (Terbeek 1977) and tones (Gandour and Harshman 1978). This paper is an attempt to discover the dimensions or features underlying the perception of Cantonese tones.

Phonological descriptions of Cantonese list six contrastive tones, that may generally be described as (1) high falling, (2) high rising, (3) mid level, (4) low falling, (5) low rising and (6) low level. Compare the following proposed representations of the Cantonese tones, in Chao (1930) tone-number notation:

* The research for this paper was supported in part by a post-doctoral fellowship (1976-1977) at Bell Laboratories, Murray Hill, N.J. I want to especially thank Osamu Fujimura of Bell Laboratories for making his research facilities available.

	1	2	3	4	5	6
Chao (1947)	53	35	33	21	23	22
Kao (1971)	53	35	33	21	23	22
Hashimoto (1972)	53	35	44	21	24	33
Vance (1976)	55	35	33	11	13	22

Of these impressionistic accounts, Hashimoto's (1972) measurements of actual fundamental frequency contours on citation forms appear to reflect the Chao (1947) and Kao (1971) tone descriptions.

Among the three earlier experimental investigations of the perception of Cantonese tones (Fok 1974; Vance 1976, 1977), the Fok study serves as the point of departure for present study of perceptual dimensions of tones. The patterns of confusions (or misidentifications) in her listening identification tests suggest that Cantonese speakers perceive separate pieces of various tonal patterns in making their identifications. If they had perceived each tone as a unitary whole, then on making a mistake, they should have been as likely to guess one tone as any other. But this did not happen; instead, two tones were most likely to be confused if they were similar in their fundamental frequency patterns, and least likely to be confused if their fundamental frequency patterns were highly dissimilar. All this suggests that Cantonese tones are perceived in terms of separate features or dimensions relatively independent of each other.

The present study uses a multidimensional scaling model of perception to investigate the number and nature of features underlying the patterns of confusions among the six Cantonese tones. The dimensions extracted from this reanalysis of the Fok data are evaluated in terms of their perceptual and linguistic plausibility, and in terms of their implications for a more general model of speech perception.

2. METHOD

2.1. *THE INDSCAL MODEL*

The output of multidimensional scaling procedures consists of a single map, or configuration, of points - one point for each stimulus. Distances between points reflect the relative similarities among objects; that is, objects which the data indicate to be more similar are in general closer to each other in the map than are less similar pairs.

In many applications of multidimensional scaling in the behavioural sciences, the similarities are obtained from several different subjects, or from the same subjects on different occasions or under different experimental conditions. Recently, a new method was developed by Carroll

and Chang (1970), and implemented in a computer program called INDSCAL (for INdividual Differences SCALing), that determines the common dimensions underlying the similarities data from different subjects or other kinds of data sources, and further determines the relative importance or weight of each dimension to every subject.

The input to INDSCAL consists of many different matrices of similarities or dissimilarities, all pertaining to the same stimulus objects. Each matrix typically comes from one person, but it is also possible for it to be associated with one of several different experimental conditions, measures of similarity, time periods, or locations. As in other multidimensional scaling procedures, the output from INDSCAL includes a map in which each point represents one stimulus object (referred to as the **group stimulus space**), but unlike other multidimensional scaling procedures, the INDSCAL output also includes a set of dimension weights for each subject (or some other data source) which shows the relative importance of each stimulus dimension to him. Subject weights may be plotted in a map in which each point represents one subject (referred to as the **subject space**).

In INDSCAL, as in other methods for multidimensional scaling, experimentation is required to determine the number of dimensions that are needed. For any specified dimensionality INDSCAL determines the stimulus co-ordinates, the subject weights, and the unique **orientation of axes** that account for the maximum total variance in the similarities data from all subjects. The distances between the stimulus objects in some latent psychological space depend on the subjects' dimension weights as well as on the stimulus co-ordinates. The program finds the particular orientation of axes that maximises the goodness-of-fit measure; in most cases, these axes or dimensions can be interpreted **without notation.** The unrotated dimensions have a special status in INDSCAL, and might be assumed to correspond to fundamental psychological process that have different saliences for different individuals or under different experimental conditions.

2.2. *FOK'S (1974) DATA ON PERCEPTUAL CONFUSIONS AMONG CANTONESE TONES*

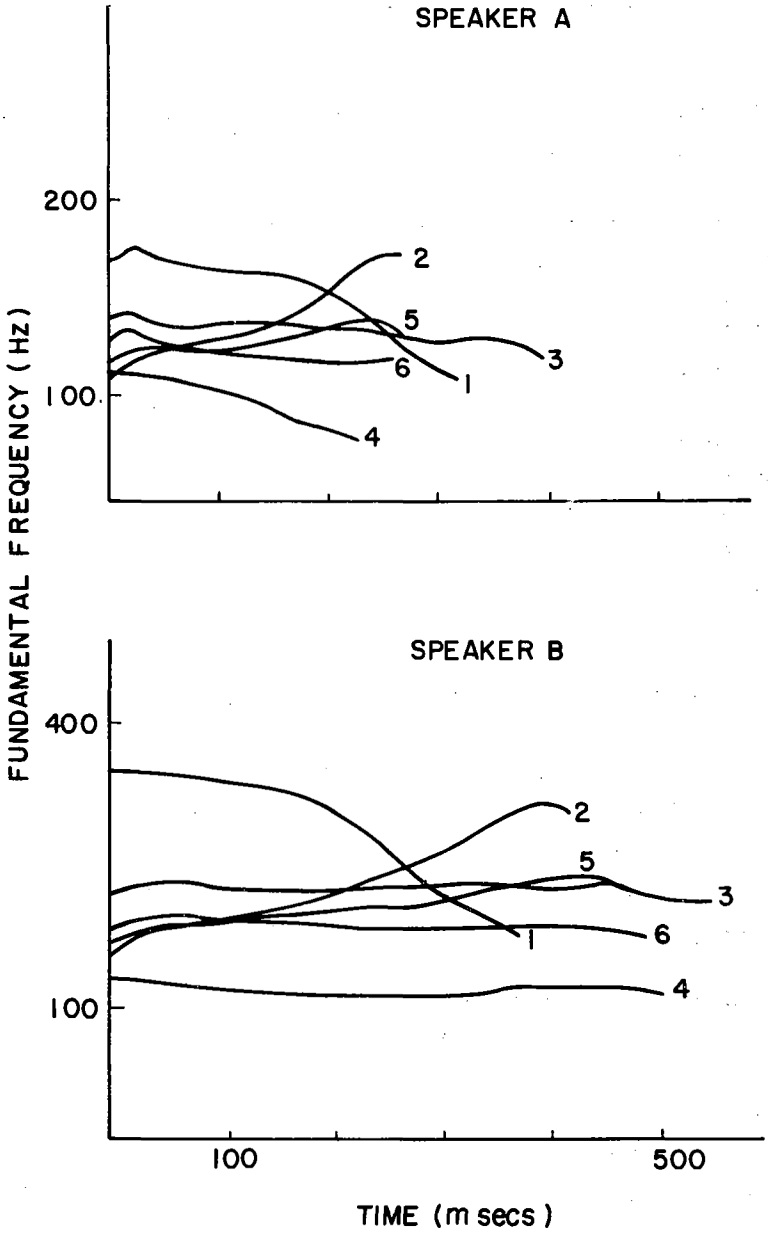
The data for this study are from an experimental investigation of perceptual confusions among Cantonese tones under different experimental conditions (Fok 1974). In one of the experiments, a male speaker (Speaker A) read the following set of words: /fu^1/ *'man'*, /fu^2/ *'bitter'*, /fu^3/ *'richness'*, /fu^4/ *'to help'*, /fu^5/ *'woman'* and /fu^6/ *'father'* from long randomised stimulus lists. Under the first experimental condition, the stimuli were simply natural speech versions of

the above set of words, produced at normal tempo with neutral mode of expression. Under the second experimental condition, the stimuli consisted of the natural larynx tones associated with this set of words. These larynx tones were obtained from direct recordings of changes in electrical impedance in the region of the larynx which occur during speech production. Under the third experimental condition, the stimuli consisted of low-pass filtered synthetic versions of these larynx tones. For each of these experimental conditions, the subjects were asked to identify the tone of the stimulus items by circling one of the words in the above set written in Chinese characters. The same experiment was repeated with a female speaker (Speaker B) reading the stimulus set. For detailed discussion of experimental method and procedure, see Fok 1974.

Figure 1 presents the fundamental frequency curves from which the various tone stimulus sets were constructed for both speaker A and speaker B. These two sets of stimuli represent two different sets of stimulus objects, and consequently must be treated separately for the purposes of a multidimensional scaling anaylsis.

FIGURE 1

Tracings of fundamental frequency curves for the six Cantonese tones (adapted from Fok 1974, by permission of Centre of Asian Studies, University of Hong Kong).



In a comparison of the two stimulus sets, it is to be noted that the speaker A tones generally display shorter duration and a narrower range of fundamental frequency, that the fundamental frequency interval between speaker A tone (1) and his remaining tones is much narrower than the difference in pitch height between speaker B tone (1) and her remaining tones, and that the fundamental frequency shape of speaker A tone (4) displays more of a falling contour than its speaker B counterpart.

A total of 511 subjects participated in the experiments, all native speakers of the Cantonese spoken in Hong Kong. There were three principal groups: 139 secondary school boys, 157 secondary school girls and 215 post-secondary school students.

Eighteen matrices of stimulus-response confusions, or errors of identification (the diagonal values indicated frequencies of correct identifications) were derived from the subjects' judgements, one for each condition for each of the three subject groups. For the speaker A stimulus set, there were nine confusion matrices (3 subject groups x 3 experimental conditions); and similarly, for the speaker B stimulus set, there were nine confusion matrices.

The entire set of eighteen confusion matrices from the Fok study was converted into similarity matrices by using the following equation (Shepard 1972):

$$\text{Sim}(i,j) = \frac{f(i,j) + f(j,i)}{f(i,i) + f(j,j)}$$

where Sim (i,j) = similarity between i and j
f(i,j) = number of confusions between i and j
f(j,i) = number of confusions between j and i
f(i,i) = number of correct responses for i
f(j,j) = number of correct responses for j

This procedure symmetrised the matrices, yielding a single value to represent the similarity associated with each pair of stimulus tones. By dividing the number of confusions between a pair of tones by the number of correct responses for that pair of tones removes the effect of bias from the resulting values.

2.3. *ANALYSIS*

Two separate INDSCAL analysis were performed - one on the speaker A stimulus set, the other on the speaker B stimulus set. The input to the INDSCAL analysis of the speaker A stimulus set consisted of nine symmetric (6 stimulus tones x 6 stimulus tones) similarity matrices.

The nine matrices represented perceptual judgements of the Cantonese tones for each of the three subject groups (secondary-level male students, secondary-level female students, post-secondary-level male and female students) across three different experimental conditions (natural speech tones, natural larynx tones, synthetic larynx tones). Each matrix contained similarity estimates for each pair of Cantonese tones for a single condition for a single subject group. The input to the INDSCAL analysis of the Cantonese tones of speaker B was similarly constructed. INDSCAL analyses of these similarity matrices were performed at several dimensionalities in order to determine the appropriate number of dimensions underlying the subjects' perceptual judgements.

3. RESULTS AND DISCUSSION

3.1. *NUMBER OF DIMENSIONS*

For INDSCAL analyses of both the speaker A and speaker B stimulus sets, three dimensions were found to provide the best representation of the perceptual structures underlying the subjects' similarities data. For the tones of speaker A, three dimensions accounted for approximately 96% of the total variance, roughly 22% more variance than two dimensions and only 2% less variance than four dimensions; for the tones of speaker B, three dimensions accounted for approximately 94% of the total variance, about 19% more variance than two dimensions and only 14% less variance than four dimensions. These results provide evidence that the three dimensions are both necessary and sufficient to characterise the subjects' patterns of confusions among the Cantonese tones.

3.2. *INTERPRETATION OF DIMENSIONS: SPEAKER A*

Plots of the first and second dimensions, and the second and third dimensions from the 3-dimensional INDSCAL group stimulus space of the Cantonese tones for speaker A are shown in the upper-half of Figure 2 (cf. Figure 1).

The first dimension is interpreted as CONTOUR. This dimension separates those tones that have a relatively steady fundamental frequency trajectory (3, 6) from those tones that show considerable changes in fundamental frequency (2, 4, 5). The position of tone (1) would appear to be inconsistent with this interpretation until one takes into account the variant shapes this tone assumes in different phonetic contexts. Tone (1), the high falling tone, becomes high level before another high falling or high level tone (Chao 1947, Huang 1965, Kao 1971), or in Chao tone-letter notation, 53 → 55/ — $\{^{53}_{55}\}$. This tone

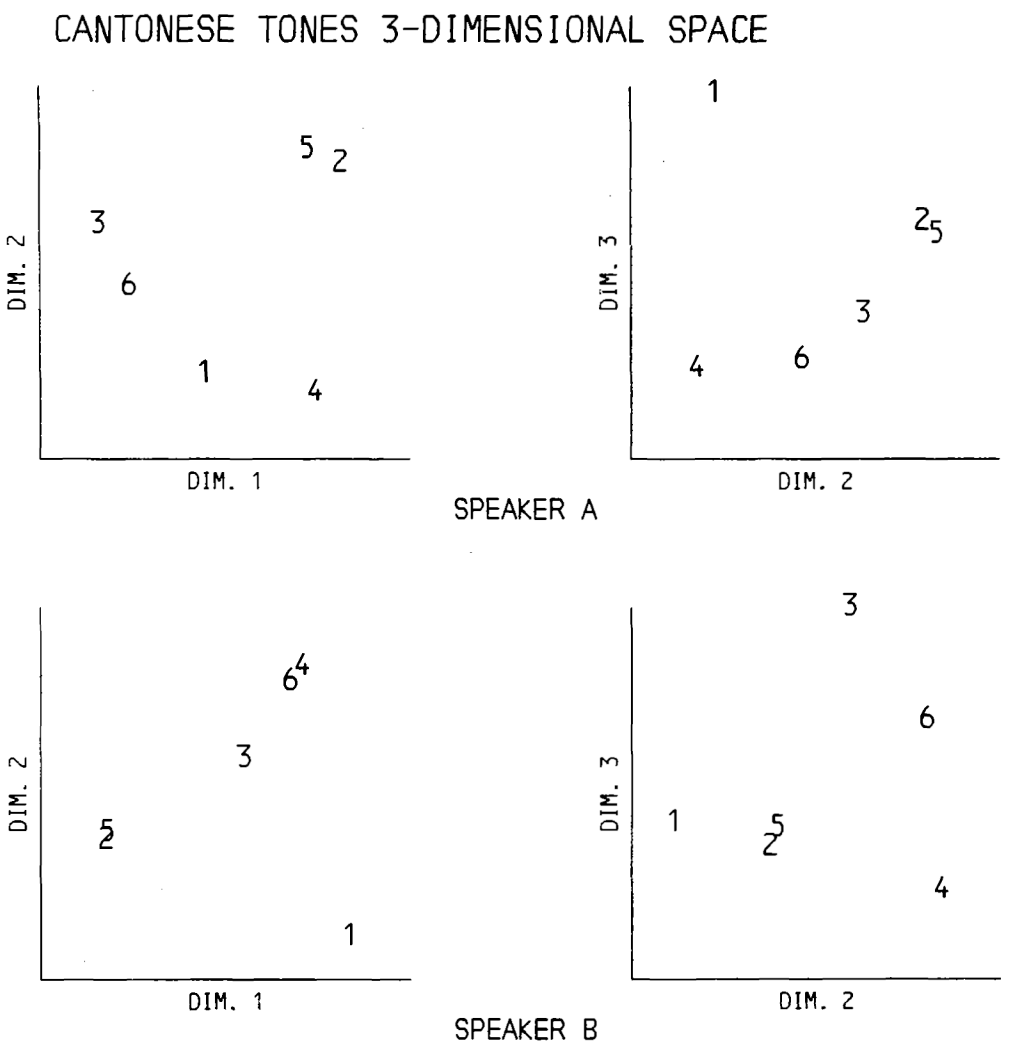
sandhi rule thus establishes a phonological relationship between a contour-shaped and flat-shaped variant of tone (1). It is this phonological relationship that accounts for the intermediate position of tone (1) on this first dimension. Despite the fact that the stimulus version of tone (1) displayed a high falling fundamental frequency pattern, the high level fundamental frequency pattern of the sandhi variant, not present in the stimulus set actually used in the Fok experiment, influenced the subjects' perception of the Cantonese tones. This result is consistent with the findings of the Vance (1977) perceptual study of the six Cantonese tones. The synthetic stimulus tones in his experiment most consistently identified as tone (1) were either high falling or high level.

The second dimension is interpretively labelled DIRECTION. This dimension places the two rising tones (2, 5) and two falling tones (1, 4) at opposite ends of this axis, and the two level tones (3, 6) in the middle. The label HEIGHT is assigned to the third dimension since the order and position of the tones on this dimension appear to be determined by average pitch height. The fact that tone (1) and tone (4) are at opposite ends of the axis as well as the fact that tone (3) falls somewhat near the middle lends support to this interpretation. The position of the tones on the third dimension indicate more crowding in the lower end of the fundamental frequency range. This result is in agreement with earlier impressionistic judgements of the Cantonese tones (Chao 1946, Vance 1977).

In this 3-dimensional INDSCAL group stimulus space, the first, second and third dimensions account for 38%, 36% and 23% of the total variance, respectively. The proportion of variance accounted for by each of the three dimensions indicates that the first and second dimensions are almost identical in saliency while the third dimension is a little weaker than the first two dimensions. The saliency of the first and second dimensions clearly reinforces the notion that, underlying the perceptual judgements of the six contrastive tones of Cantonese, speakers make decisions about whether the tone is level or contour in shape and whether the tone is level, rising or falling in direction.

FIGURE 2

Dimensions 1 and 2 (upper left, lower left) and dimensions 2 and 3 (upper right, lower right) of the group stimulus space from three-dimensional INDSCAL analyses of the Fok tone confusability data for speaker A and speaker B, respectively.



3.3. *INTERPRETATION OF DIMENSIONS: SPEAKER B*

Plots of the first and second dimensions, and the second and third dimensions from the 3-dimensional INDSCAL group stimulus space of the Cantonese tones for speaker B are shown in the lower half of Figure 2 (cf. Figure 1).

The first dimension is interpretively labelled DIRECTION. This dimension places tone (1) and tones (2, 5) at opposite ends of the axis, with tones (3, 4, 6) in the middle. It is nearly identical to the second dimension for the speaker A perceptual space except for the position of tone (4). In the speaker B perceptual space, tone (4) clusters with tones (3, 6); in the speaker A perceptual space, tone (4) clusters with tone (1). This difference in the positioning of tone (4) on the DIRECTION dimension can be understood in terms of the actual shape of tone (4) in the speaker A stimulus set and the speaker B stimulus set, respectively. In the former, tone (4) displays a falling fundamental frequency pattern; in the latter, tone (4) displays a relatively level fundamental frequency pattern. Such realignment of different stimulus tones along what appears to be the same dimension provides further confirmation of its psychological reality.

The second dimension, similar to the first dimension of the speaker A perceptual space, separates the level tones from the contour tones - this is, tones (3, 6, 4) from tones (1, 2, 5). This dimension is labelled CONTOUR. Unlike the first dimension in the speaker A perceptual space, however, tone (1) is not perceived to be similar to the level tones, perhaps because of the much wider pitch interval between tone (1) and tones (3, 4, 6) in the speaker B stimulus set.

The third dimension in the speaker B group stimulus space, while the most difficult to interpret, can be interpreted as a HEIGHT dimension. The order and position of tones (3, 6, 4) along this dimension closely correspond to the actual pitch height of the stimulus versions of these tones. Thus, tone (6) is perceived to be more similar to tone (3) than tone (4). The problem lies with the contour tones (1, 2, 5). These three tones were perceived to be virtually identical along this dimension. Why they cluster toward the lower end of the fundamental frequency range remains unclear.

In this 3-dimensional INDSCAL group stimulus space for the Cantonese tones of speaker B, the first, second and third dimensions account for 40%, 34% and 20% of the approximate total variance, respectively. For both the speaker A and speaker B stimulus sets, the CONTOUR and DIRECTION dimensions together accounted for 74% of the variance. Their relative importance, however, varies across the stimulus sets, with the

DIRECTION dimension being more predominant in the perceptual judgements of the Cantonese tones of speaker B. The third dimension, HEIGHT, turns out to be the weakest of the three dimensions in this experimental task, for both the speaker A and speaker B sets of Cantonese tones. The emergence of three interpretable dimensions in the two separate INDSCAL analyses, similar in the order and position of the tones along the dimensions and plausible from both a perceptual and linguistic standpoint, confirms the validity of the multidimensional scaling alanysis. It is surely not accidental that the 'same' three dimensions underlie these two slightly different sets of Cantonese tones, and that the relative importance of these three dimensions to each other depends on the exact composition of the stimulus tones.

3.4. *DIMENSION WEIGHTS FOR EXPERIMENTAL CONDITIONS*

In general, there was little systematic variation in the patterns of dimension weights for the three different kinds of stimuli - natural speech, natural larynx tones and synthetic larynx tones. This would seem to indicate that pitch variation is the principal cue for distinguishing the six Cantonese tones. In other experimental investigations of Cantonese tones, Fok (1974) and Vance (1976) have shown that concomitant differences in syllable duration are not important for Cantonese tonal distinctions.

For the speaker A set of Cantonese tones, however, weights on the CONTOUR dimension are highest under the natural speech condition; for the speaker B set of Cantonese tones, weights on the HEIGHT dimension are highest under the natural speech condition. This perhaps reflects the differences in degree of movement in fundamental frequency patterns between the two stimulus sets. Under the natural speech condition, subjects optimally directed their attention to these general properties of the two sets of tones. While this is not a wholly satisfactory explanation, nevertheless, these results do show that the dimensions obtained and their relative importance may depend on the experimental conditions as well as the stimuli and subjects. In the Fok study, no differences in the perception of Cantonese tones could be attributed to the three different subject groups. This result is demonstrated in the INDSCAL analyses in the lack of differences in the patterns of dimension weights across subject groups.

4. FURTHER DISCUSSION

The particular three dimensions that emerged from the INDSCAL analyses of the Cantonese tone confusions - CONTOUR, DIRECTION and

HEIGHT - are meaningfully interpretable in both perceptual and linguistic (cf. Wang's (1967) proposed set of phonological features of tone) terms. This is important for providing us a better understanding of how Cantonese speakers construct an internal representation of the tonal patterns. The fact that we obtained perceptually and linguistically plausible dimensions for Cantonese tones takes on even greater significance when compared to the results of other multidimensional scaling investigations of tone perception. Gandour and Harshman (1978), in their cross-language study of tone perception, which included two typologically and genetically unrelated tone languages (Thai, the national language of Thailand, and Yoruba, a language spoken primarily in Nigeria), also found dimensions that could be similarly labelled. In another multidimensional scaling analysis of the perception of tones in Yoruba, Homhert (1976) found dimensions that could be related to the direction of movement in fundamental frequency, and the distinction between level and contour fundamental frequency shapes. The present study as well as these two earlier ones employed different stimuli, subjects and experimental tasks; yet the dimensions extracted are similar enough to suggest that such dimensions or features must, indeed, be psychologically real, and part of the universal set of phonetic/phonological features underlying the perception of tone. The precise number and nature of these features, of course, is a subject for future experimental investigation.

Linguists are not only interested in the construction of a universal set of phonetic/phonological features, but they also wish to know to what extent these features are utilised in particular languages. Or to put it in a different perspective, how does the language background of a listener affect his perception of speech sounds; in the context of this study, does the structure of Cantonese influence his perception of the six contrastive tones? Although our data lacks a control group for comparison, the position of tone (1) on the CONTOUR dimension in the speaker A group stimulus space strongly suggests that phonological rules may influence a person's perception of speech sounds. This effect of particular linguistic experience is also shown in the Gandour and Harshman (1978) study, where differences in the composition of tonal inventories, or the lack thereof, is reflected in differential emphasis placed on selected dimensions by speakers of Thai, Yoruba and English. Such data point to the need for a model of speech perception that incorporates higher-level linguistic information into the perceptual processing of speech signals.

And finally, the application of INDSCAL has proven to be a very useful tool for learning a great deal about both the stimuli and the subjects under investigation, for discovering the underlying dimensions of a multidimensional perceptual space, and for confirming and improving on specific hypotheses concerning human perception. These findings can now be incorporated into the design of other laboratory experiments dealing with the processing of tone as well as other speech sounds.

5. SUMMARY

A multidimensional scaling reanalysis of the Fok (1974) tone confusion data from Cantonese revealed three underlying perceptual dimensions that were interpretively labelled CONTOUR, DIRECTION and HEIGHT, respectively. The influence of a Cantonese tone sandhi rule on a listener's tonal perception was evident in the position of the high-falling tone (1) on the CONTOUR dimension. The weights of these dimensions were not found to vary much across the different kinds of tonal stimuli - natural speech tones, natural larynx tones and synthetic larynx tones. These dimensions extracted from the patterns of tonal confusions in Cantonese were found to bear close resemblance to perceptual dimensions of tone that have emerged in other multidimensional scaling investigations.

BIBLIOGRAPHY

CARROLL, J.D. and J.J. CHANG

1970 'Analysis of Individual Differences in Multidimensional Scaling via an n-way Generalization of "Eckart-Young" Decomposition'. *Psychometrika* 35/3:283-319.

CHAO, Y.R.

1930 'A System of "Tone Letters"'. *La Maître Phonetique* 32: 24-7.

1947 *Cantonese Primer.* Cambridge, Mass.: Harvard University Press.

FOK CHAN YUEN-YUEN

1974 *A Perceptual Study of Tones in Cantonese.* Centre of Asian Studies Occasional Papers and Monographs, 18. Hong Kong: Centre of Asian Studies, University of Hong Kong.

GANDOUR, J.T. and R.A. HARSHMAN

1978 'Crosslanguage Differences in Tone Perception: A Multi-dimensional Scaling Investigation'. *Language and Speech* 21/1:1-33.

HASHIMOTO, O-K.Y.

1972 *Studies in Yue Dialects* 1: *Phonology of Cantonese.* London: Cambridge University Press.

HOMBERT, J.M.

1976 'Perception of Tones of Bisyllabic Nouns in Yoruba'. *Studies in African Linguistics, Supplement 6*, 109-21.

HUANG, P.
1965 *Cantonese Sounds and Tones*. New Haven, Conn.: Yale University, for Eastern Publications.

KAO, D.L.
1971 *Structure of the Syllable in Cantonese*. The Hague: Mouton.

SHEPARD, R.N.
1972 'Psychological Representation of Speech Sounds'. In: E.E. David and P.B. Denes, eds *Human Communication: A Unified View*, 67-113. New York: McGraw-Hill.

SINGH, S.
1975 'Distinctive Features: A Measurement of Consonant Perception'. In: S. Singh, ed. *Measurement Procedures in Speech, Hearing and Language*, 93-155. Baltimore: University Park Press.

STUDDERT-KENNEDY, M.
1975 'Speech Perception'. In: N.J. Lass, ed. *Contemporary Issues in Experimental Phonetics*, 243-93. Springfield, Illinois: Charles C. Thomas.

TERBEEK, D.
1977 'A Crosslanguage Multidimensional Study of Vowel Perception'. *Working Papers in Phonetics, University of California, Los Angeles* 37.

VANCE, T.J.
1976 'An Experimental Investigation of Tone and Intonation in Cantonese'. *Phonetica* 33/5:368-92.

1977 'Tonal Distinctions in Cantonese'. *Phonetica* 34/2:93-107.

WANG, W.S-Y.
1967 'Phonological Features of Tone'. *International Journal of American Linguistics* 33/2:93-105.