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 charcter-ize 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.