The Scaling of Speech Sounds

The perception of synthetic speech has been investigated mainly through experimental work on discrimination and identification functions. Scaling has been little used, but as a research tool, its potential in this field is great. In other areas of research, scaling studies have provided quantitative information about psychophysical phenomena in a simple and direct fashion. The procedure is usually uncomplicated and the data relatively easy to collect. Scaling techniques show where along a stimulus continuum a subject is able to detect changes, and also provide a relative measure of the phenomenal size of each change: the data are a kind of psychological mapping of the stimulus dimension.

The present investigation is mainly concerned with two problems:

(1) Can reliable scaling methods be developed which work with stimuli such as synthetic speech sounds?

(2) Will such methods be sufficiently sensitive to show up perceptual differences between different classes of speech sounds similar to those previously found in studies of the discrimination and identification of these sounds?

The stimuli in current use are synthetic speech sounds for which fairly extensive discrimination and identification data are available. This will enable the scaling data to be directly evaluated in terms of what is already known.

Many scaling methods are to be found in the psychophysical

literature. The method which has seemed most promising for present purposes is direct magnitude estimation. The principle of the method is as follows: a subject is presented with a series of stimuli and responds by assigning numbers, one number to each stimulus. If the subject has been given suitable instructions, the numbers may be assumed to represent the relative psychological magnitudes of the stimulus differences. The reliability and usefulness of this method are documented in the psychological literature. Instead of assigning numbers, a subject may draw lines of appropriate length or mark off lengths on a line drawn for him.

In the present study, preliminary data suggest that the method has the same order of reliability as found with non-speech stimuli.

The stimuli are presented in randomized pairs and the subject responds to the difference he perceives between the two sounds in the pair. A truncated version of this experiment has shown that there are marked differences in the way that subjects respond to steady state vowels and stops.

For vowels, the plot of psychological magnitude against the stimulus dimension gives an essentially smooth, continuous function. For stops, sharp breaks occur, dividing the curve into distinct and almost flat segments. The perception of stops would seem to be categorical, that of vowels, continuous. Fuller experiments, however, are needed to verify this result. Finally, it is hoped that vowels in dynamic context may be introduced so that direct comparisons of the mode of perception for these three classes of synthetic speech sound may be made.

Michael Vinegrad