Usefulness of the Mean in Psychological Scaling of Cleft Palate Speech

DOROTHY SHERMAN, Ph.D.

Iowa City, Iowa

Psychological scaling techniques have frequently been used for research purposes of obtaining measures of various aspects of speech of individuals with cleft palate (*3*). These techniques all require judgments by observers, or listeners, for each of an array of stimuli, that is, in this case, an array of speech samples. Techniques which have been used to obtain measures for cleft palate speech samples include the method of pair comparisons, the method of direct magnitude-estimation, and the method of equal-appearing intervals. Of these three methods the most frequently used has been the method of equal-appearing intervals (*3*).

Application of the method of equal-appearing intervals (*2, 7*) includes rating by observers, or listeners, of each of an array of speech samples on a scale under the assumption that the interval between any two adjacent points on the scale is equal to the interval between any other two adjacent points on the same scale. Often used is a *seven*-point scale extending from *one*, for least, to *seven*, for most, of the particular aspect of the stimuli to be rated.

The present purpose is to evaluate the usefulness of one of two possible methods of computing equal-appearing-intervals scale values for samples of cleft palate speech for each of the following: nasality, articulation, language development, and general defectiveness.

Scale values obtained by the method of equal-appearing intervals have customarily been the medians of the sets of ratings assigned by the observers to the stimuli (*2*). Use of the medians rather than the means avoids undesirable influence upon the scale values from possible skewing of distributions of ratings. That is, extreme ratings toward one end of the scale which are far from close agreement with the other ratings would have less effect upon the median than upon the mean of the set of ratings.

The extremes of the scale, that is, *one* and *seven* on a *seven*-point scale, or *one* and *nine* on a *nine*-point scale, are identified by the extremes in the particular array of stimuli for which scale values are obtained. The points on a scale for one array of stimuli thus cannot be assumed to have the same value as the same points on a scale for another array of stimuli.

A point has been made, in a report on another experiment (6), that use of means rather than medians for equal-appearing-intervals scale values, when this is feasible, is advantageous particularly when computer service is available. An analysis of variance designed for repeated measurements on the same subjects will provide not only the desired mean scale values but also the information needed to compute a coefficient for estimating the reliability, that is, the repeatability, of the mean scale values. The particular measure referred to is the intraclass correlation coefficient for estimating the reliability of means (1).

The question of whether mean equal-appearing-intervals scale values can feasibly be used instead of median equal-appearing-intervals scale values is dependent upon the closeness of the relationship between the set of means and the set of medians obtained from ratings of an array of stimuli. High Pearson rs for estimating the relationship between two such sets of measures, indicating a very close relationship, have been reported for several types of stimuli (6). These types of stimuli, however, do not include cleft palate speech.

The specific purpose of the experiment reported here has been to evaluate, by means of the Pearson r procedure, the feasibility of using mean equal-appearing-intervals scale values rather than median equal-appearing-intervals scale values as measures of the various aspects of cleft palate speech previously mentioned.

Procedure

STIMULI. Tape recordings of the speech of 154 children with cleft palate were available from a previous experiment (8). Portions of the 154 tape-recorded speech samples, each five seconds long, were the stimuli which were rated on a *seven*-point equal-appearing-intervals scale with reference to a) degree of nasality, b) defectiveness of articulation, and c) general defectiveness.

Typed samples of the speech of 82 children with cleft palate, each consisting of approximately the first 150 words of a sample originally elicited for a previous experiment (4), were the stimuli which were rated on a *seven*-point equal-appearing-intervals scale with reference to degree of language development.

OBSERVERS. The 37 observers who rated the 154 tape-recorded speech samples for articulation defectiveness and for degree of nasality were all students majoring in speech pathology who had had some training and experience in the areas of articulation and voice. Although the same group of observers rated the same samples twice, it is unlikely that the second rating was influenced by the first because of the fact that the nasality ratings were made with the samples presented by backward play (5).

The 56 observers who rated the 154 samples for general speech defectiveness were students in an introductory speech pathology course.

The 101 observers who rated the 82 typed language samples were students in an elementary psychology course.

RATING SESSIONS. For each of the rating sessions the observers were instructed to rate each sample in relation to a *seven*-point scale of equal intervals, from *one* to *seven*, with *one* representing least and *seven* representing most of the particular aspect of cleft palate speech to be rated: nasality, defectiveness of articulation, language development, or general defectiveness. They were told that *four* is the middle point between *one* and *seven* with the other points spaced for equal distances, or intervals, between adjacent points.

To establish for the observers the range from point *one* to point *seven*, samples representative of the two extremes of the aspect of speech to be rated were presented at the beginning of each rating session. Also for the purpose of establishing the range, and for practice, observers rated 20 samples which they again rated during the experimental judging.

No more than 20 observers participated in the experiment during any one rating session. Previous experience of the experimenter had provided evidence that a maximum of 20 observers participating at any one time in psychological scaling of an array of stimuli is to be preferred for the purpose of procedural efficiency.

Results

Two sets of psychological scale values, consisting of a) medians and b) means, were obtained by the method of equal-appearing intervals for arrays of language samples from children with cleft palate for each of four aspects of speech: nasality, articulation, language development, and general defectiveness. The agreement between the two sets of scale values for each of the four aspects of speech was evaluated by means of the Pearson correlation procedure. The obtained correlation coefficients are high: for nasality the *r* is .96; and for articulation, for language development, and for general defectiveness the *r* is in each instance .99.

The close agreement between the two sets of scale values for each of the four aspects of speech indicates that a choice between mean and median equal-appearing-intervals scale values, depending upon the experimenter's purposes, or convenience, is warranted.

Summary

The agreement between two sets of psychological scale values, means and medians, obtained by the method of equal-appearing intervals was evaluated by the Pearson correlation procedure for arrays of speech samples from children with cleft palate for each of four aspects of speech: nasality, articulation, language development, and general defectiveness. The obtained Pearson *r*s of .96 for nasality and .99 for each of the other three aspects of speech indicate very close agreement between the means and the medians. The conclusion is that the experimenter who obtains

equal-appearing-intervals psychological scale values for any of these four aspects of cleft palate speech samples is justified in choosing between medians and means to suit his convenience in analyzing the data. A secondary purpose of the article is to present some specific information on the psychological scaling method of equal-appearing intervals and the scale values obtained by this method.

reprints: *Dr. Dorothy Sherman*
Wendell Johnson Speech and Hearing Center
The University of Iowa
Iowa City, Iowa 52240

References

1. EBEL, R., Estimation of the reliability of ratings. *Psychometrika, 16,* 407–424, 1951.
2. EDWARDS, A. L., *Techniques of Attitude Scale Construction.* New York: Appleton-Century-Crofts, Inc., 1957.
3. MOLL, K. L., Speech characteristics of individuals with cleft palate. *In* D. C. Spriestersbach and Dorothy Sherman (eds.), *Cleft Palate and Communication.* New York: Academic Press, 1968.
4. MORRIS, H. L., Communication skills of children with cleft lips and palates. *J. speech hearing Res., 5,* 79–95, 1962.
5. SHERMAN, DOROTHY, The merits of backward playing of connected speech in the scaling of voice quality disorders. *J. speech hearing Dis., 19,* 312–321, 1954.
6. SILVERMAN, F. H., Correspondence between mean and median scale values for sets of stimuli scaled by the method of equal-appearing intervals. *Precep. motor Skills, 25,* 727–728, 1967.
7. THURSTONE, E. L., and E. CHAVE, *The Measurement of Attitude.* Chicago: University of Chicago Press, 1929.
8. VAN DEMARK, D. R., Misarticulations and listener judgments of the speech of individuals with cleft palates. *Cleft Palate J., 1,* 232–245, 1964.