SOME DETERMINANTS OF ERROR RESPONSES IN THE AURAL IDENTIFICATION OF WORDS

JOHN W. BLACK

E.W. Scripture proposed that errors in aural identification come with associations, (1) ones of phonetic similarity and (2) ones of meaning (Scripture 1902). An object of the study reported here was to investigate the relative potency of these. Two sets of hypotheses were tested. First, errors in word perception are not identified with either associations of sound or of semantics. Second, there is no relationship between the intelligibility of a word and the relative similarity-dissimilarity of the word and listeners' error responses to the word. Additionally, a test was made of the extent to which the different intervals of a 9-point equal-appearing intervals scale is utilized in making judgments pertinent to the foregoing hypotheses (Sherman and Moodie 1957; Sherman and Silverman 1968).

1. PROCEDURE: STIMULI

The error responses that had been made by approximately 100 listeners while responding to 3500 words were available. The words were recorded by ten male speakers in quiet and heard through headsets in the presence of 110 dB of ambient noise (General Radio sound level meter, C-scale). The auditors' responses to a word were entered on a card headed by the stimulus word. These cards, on per stimulus word, were filed alphabetically and became the source of the present stimuli. One hundred cards were selected from the file. These cards were equally spaced among the ones on which there were entries of at least ten different error responses. The stimulus word and one error response were typed in a paired manner on a 3 × 5 inch file card. From ten to 40 of these cards related to a single stimulus. These were ring-bound and designated a SET. Thus, there were 100 sets of cards that comprised the stimuli.

2. SUBJECTS

Fifteen undergraduate students of phonetics worked individually in a task that required from two to five hours per subject and was completed within a two-week

period. The instructions were: how much alike is the SOUND of these two words? Do they seem very similar (easily confused) or are they quite different? Please encircle one number in each answer group. Use the number 1 to mean 'very similar' and 9 to mean 'very dissimilar' and 2, 3, 4, 5, 6, 7, and 8 to mean 'degrees' on a continuous scale from 'very similar' to 'very dissimilar'. Example: (very similar) 1 2 ③ 4 5 6 7 8 9 (very dissimilar). Explanation: Since 3 is encircled, this pair of words was judged by one person to have some but not a great amount of similarity. Please make up your own guidelines. The omission of definitions and specific examples of pairs of words is deliberate. Therefore, please do not ask anyone for help.

Another group of 15 undergraduate students of psychology in the same university followed these instructions while working with the same cards: When the first word on each card of a set of cards was heard it was identified as the second word on the card. What is the degree of association, that is of association in terms of MEANING, of the two words TO YOU? If the association is close, please consider it to be 1; if the association is remote, very distant, please consider it to be 9. Rime or other similarity in the sound of the two words has nothing to do with the judgment. You are evaluating the degree of association between the two words in YOUR EXPERIENCE. (The omission of an example is deliberate.) Both the sets and the cards of a set are numbered. The number of the set should be entered one time on a page (answer sheet). The number of the card within a set should correspond with the number of an answer-group. Here is an example of an answer group:

(close association) 1 2 3 4 5 6 7 8 9 (no association)

3. DATA

The ratings of the 15 judges on each of the two equal-appearing interval scales were pooled for each STIMULUS WORD vs. each ERROR RESPONSE WORD. A hypothetical example follows (it would not have been used, for it accounts for only seven different error responses, not ten): (1) of 97 responses 57 were correct (there were 40 errors), (2) one word accounted for 19 of 40 errors, (3) another word accounted for 9 of 40 errors, (4) a third word accounted for 6 of 40 errors, (5) a fourth word accounted for three of 40 errors, and three other words accounted for three of 40 errors, one each. In this summary the rankings 'first most-frequent error', 'second most-frequent error', and the like apply with (6) or 'the single instances' pooled as the 'fifth most-frequent error'. To extend the example, the mean values that 15 judges assigned the words of (2), (3), (4), (5), and (6) on an equal-appearing interval scale might be 1.3, 2.1, 3.1, 4.0 and 6.2 respectively. Weighting these values by the frequencies with which (2), (3), (4), (5) and (6) occurred would yield a mean for the 40 error responses of 2.3 on a scale 1-9. This would represent the distance of the pooled judgments of the error responses from the single stimulus word. On another dimension,

the intelligibility of the word would be the ratio of the number of correct identifications to the number of responses, or 58.8 %.

4. RESULTS

Three aspects of the relation between the responses and the stimuli were under scrutiny: (1) the position on a scale of similarity-dissimilarity assigned to the sound of error-words relative to the words the listeners were intended to hear; (2) the position on the same scale of the values assigned in terms of meaning; and (3) the relation between these two distances and the intelligibility or correct identifications of the words.

There were 1561 responses made in error to the 100 stimulus words. The mean value on a 1-9 scale assigned to these 100 sets of words with respect to similarity-dissimilarity in sound was 6.2 ± 0.8 (range: 4.1-8.3). This contrasts with values assigned in terms of meaning or definition, mean 7.4 ± 0.9 (range 4.8-8.4). Thus, the errors are more like the stimuli in sound than in meaning.

The stimuli varied in intelligibility from 12-87 %, from 4.1-8.3 when judged on the basis of sound. The correlation coefficient between the 100 intelligibility scores and the 100 values representing the weighted indexes of similarity-dissimilarity was, r, -.21, statistically significant. This suggests a tendency for the errors associated with highly intelligible words to sound more like the stimulus word than in the instance of less intelligible words.

Another approach was undertaken to the matter of the acoustic similarity of the stimuli and the error responses to the stimuli. The responses were dichotomized between MULTIPLE and SINGLE responses, that is between words that were written by more than one listener and ones that were written by only one. The means of the similarity-dissimilarity ratings for the single responses (6.0) and the weighted means of the multiple responses (4.7) were computed for each of the 100 sets (t, $13.98 < 1$ %).[1] Obviously the error responses in a task of speech perception have some similarity in sound to the stimuli, rejecting a portion of the first hypothesis.

In connection with the second portion of the first hypothesis, the relation between the stimulus and the response in meaning approached the upper end of a 9-point scale ($7.4 \pm .09$, range 1.6-8.4). The need for defining the upper pole of a 9-point equal-appearing interval scale in the task seemed clear, that is to test whether or not 7.4 was the upper end of the scale. To examine the topic, the 100 sets of words were re-ordered. Stimulus word 1 was paired with the responses to stimulus word 10; stimulus word 2, with the responses to stimulus word 11, etc., with the stimulus word 100

[1] A confirming approach to this result was made. This was limited to the initial portions of the words. The relative perceptual distances between the initial consonantal phonemes and clusters of English words have been obtained. These values were applied to the single- and the multiple-error responses and the two distributions were compared (t, $3.51 < 5$%).

paired with the responses to word 9. The pairs of words had no common origin. The procedures of judging were applied a second time with different panels of judges selected and instructed as the first ones. The artificially created pairs yielded a mean for similarity in SOUND, 8.0 ± 0.4 (range, 6.3-8.6); for similarity in MEANING, 7.7 ± 0.2 (range, 7.0-8.2).

With no reason to assume a relation between the 1561 pairs of words, 7.7-8.0 might be considered to be an upper pole for a 9-point scale under the conditions of the study. A comparison of the means that pertained to semantics, 7.4 and 7.7, was made with scores associated with each stimulus word, (N, 100) entered in paired columns (t, $7.5 < 1$ %). This prompted an examination of the words that accounted for ratings that showed high semantic associations (low scale values) in the original comparison. 74 of the 1561 values were isolated as accounting for the skewed distribution. Preponderantly, these pairs of words had derivations and roots in common, e.g., active-acted, avenge-revenge, brute-bruise, bother-bothered, cherub-cherished, and brim-rim. The few exceptions included abate-evade, boat-hope, herdsman-person. With the values of pairs of words of common derivation removed from the distribution, the two sets of scores did not differ.

5. DISCUSSION AND CONCLUSION

The results demonstrated that acoustic association is somewhat present in the errors that occur in speech perception. Scripture's (1902) further attribution of these errors to semantic association was also borne out to the extent that errors in aural perception of speech may have roots and derivations in common with the stimulus words. Thus both portions of the hypothesis that prompted the study were rejected. Likewise the hypothesis that there is no relationship between the intelligibility of a word and the similarity in sound between it and the error responses that are made to it was rejected.

Another outcome of the procedures of the study relates to the 9-point scale. Although a single judge might use the entire scale individual differences would reduce the range of the obtained mean values assigned to stimuli. With 15 judges and tasks that might be expected to 'exhaust' the range, individual scores accounted for a 9-point range and mean scores for a 5-point range. The upper limit in the rating of haphazardly paired stimuli was of an order 7.7-8.0.

In spite of the foregoing positive outcome of the study, the relatively high magnitude of the weighted mean of values for sound may be viewed as a somewhat negative result. The hypothetical example led to a score of 2.3. This anticipated result was lower than the score for any single set and bespeaks the experimenter's surprise. First, a contaminating factor was the requirement that only words would be considered with at least 10 different error responses. This exceeds the available number of responses of rhyming words and the like and assures some responses of limited

relationship in sound to the stimulus. Second, the original listening was in noise. This probably contributed variability to the responses.

Overall, Scripture was right. His 'arm-chair' statement may have been more correct than the data presented here would suggest.

Speech Department
Ohio State University
Columbus, Ohio

REFERENCES

E. W. Scripture
1902 *Elements of Experimental Phonetics* (New York, C. Scribner's).
Sherman, D. and C.E. Moodie
1957 "Four Psychological Scaling Methods Applied to Articulation Defectiveness", *Journal of Speech and Hearing Disorders* 22:698-706.
Sherman, D. and F. Silverman
1968 "Three Psychological Scaling Methods Applied to Language Development", *Journal of Speech and Hearing Research* 11:837-841.

DISCUSSION

HARMS (Honolulu)
At an earlier time, you reported on the success of a speaker in hearing himself after a slight delay. Have you continued this line of research?

BLACK
This study is unrelated to the series to which you refer. As for that series, which is not currently being continued, no evidence that a person could understand himself better than others emerged.

SINGH (Washington)
Did you control for the graphemic similarity?

BLACK
The method of the study forbade the control of spelling. Further, since the results show a positive relation of the stimulus and response in SOUND and since there is an inevitable relationship between spelling and sound, it follows that some relation was present between the obtained judgments in SOUND and the spelling of the stimulus and the response.