Multiple Choice Speech Discrimination Tests for Both Diagnostic and Rehabilitative Evaluation: English and Spanish

J. C. COOPER, Ph.D.
and
LYNN B. LANGLEY, M.A.
Department of Physical Medicine and Rehabilitation
and
Division of Otolaryngology
Department of Surgery
The University of Texas Health Science Center
at San Antonio

The Modified Rhyme Test (MRT) and the Spanish Multiple Choice Rhyme Test (SMRT) were used as materials in a normative study of auditory, visual and audiovisual performance. Monaural presentations in three levels of noise were made to 60 English and 60 Spanish speaking, normal hearing subjects. Results suggest that the MRT can be adapted for both diagnostic and rehabilitative evaluations. The SMRT appears to be satisfactory for diagnostic evaluation although further exploration of its audiovisual performance is warranted.

A simple listing of the tests available for evaluating the hearing impaired would be staggering. It is our impression that the limited number of tests actually used with any given individual is as much a function of available time as of anything else. Particularly during an initial evaluation, measures which focus on rehabilitative aspects are likely to be curtailed. Further, the available tests are not readily comparable to those used in the diagnostic phase. This report is an attempt to standardize speech discrimination materials which can be employed for both diag-
nostic and rehabilitative purposes and, by yielding comparable scores, have the potential for increasing the efficiency of audiological practice.

Before turning to particulars, it is appropriate to briefly review the factors affecting overall communication in general and the antecedents of our specific approach. The large number of factors involved in communication may be roughly sorted into four categories:

1. the message, ranging from descriptors of the phonetic content through to the broadest meaning of communication context,
2. the transmission system, from its frequency response through to its signal-to-noise ratio,
3. the listener, from considerations of pathophysiology through to central factors such as his familiarity with the message, and
4. the examiner, particularly his ability to unambiguously interpret the listener’s response.

The wide variety of factors precludes any practical examination which would explore all the relevant effects. Compromises must be made, though their nature should be made explicit. In this study they include:

1. the use of words lists, which for rehabilitative purposes are clearly atypical of normal communication situations in that contextual information is absent,
2. the use of a “multiple-guess” format to re-establish the limited response set more characteristic of contextual cues, and to reduce examiner variability,
3. the use of a recorded format to stabilize the test stimuli (if at the expense of reduced flexibility),
4. the use of standard speech noise and low sensation levels to more closely approximate real life communication environments and to permit independent replication of materials, understanding that competing speech or cocktail party noise would be a better approximation,
5. the use of monaural presentation to permit evaluation of each ear in spite of the clearly documented advantage derived from binaural hearing, particularly in noise and
6. the availability of Spanish language materials which approximate the English since our particular community has a high proportion of individuals whose native tongue is intimately related to Spanish, even though not identical.

The selection of particular word lists was governed by the dual purpose sought for the materials. The most readily adaptable appeared to be the Modified Rhyme Test (MRT) (Krueel et al., 1968) and what, by reasonable translation, we have called the Spanish Multiple Choice Rhyme Test (SMRT) (Tosi, 1969). In the realm of auditory-only presentation, previous reports (Kopra, et al., 1968 and Lovrinic, et al., 1968) suggest that
the MRT is a reasonable measure of discrimination, if not the most
discerning. A particular advantage to the MRT is its apparent resistance
to learning effects (House, et al., 1965), since in its proposed application
it would be appropriate to evaluate an individual's performance at se-
lected intervals during rehabilitative progress as well as recommending at
least two measures of function during the initial evaluation. It approxi-
mates the performance-intensity function of CID W-22 with a variety of
subjects and, if easier (at least because of the closed response set) than
recordings of lists from Harvard's Psychoacoustic Laboratory, certainly
open to no more criticism than could be leveled at variations of discrimi-
nation materials in daily use. It was judged that the MRT would be
satisfactory when employed with noise sufficient to reduce a normal
hearing individual's score below 90%. Unfortunately, reports of the
replication of initial findings with the SMRT could not be uncovered.

Although the auditory functions of both tests were available, no report
could be found of their analogous visibility functions. From the many
reports of improved performance with audio-visual presentation of word
lists, it becomes evident that most word lists are not adaptable for lipread-
ing alone. The CID W-22 lists result in only some 13% mean lip reada-
ability (Sanders and Goorrich, 1971) and the CNC lists of Lohiste result
in only 34% (Duffy, 1967). A variation of the multiple-guess format list is
reported to produce some 60% correct identification at a -20 dB signal-to-
noise ratio (Neely, 1956).

Taken together, we believe that previous reports supported the con-
clusion that the MRT and SMRT provide the best opportunity to amalgam-
ate the demands of both diagnostic and rehabilitative evaluations and
that they provide the advantage sought initially: a single set of materials
which can be used interchangeably.

SUBJECTS

One hundred twenty adults with hearing at or better than 20 dB HTL
(ANSI, 1969) at octave intervals from 250-4000 Hz in both ears partici-
pated in the study. Sixty were native speakers of general American dialect
and sixty were native Spanish speakers. No differentiation was made
among Spanish speaking subjects on the basis of the geographical region
or socioeconomic status associated with their language acquisition. This
issue has been raised in relation to such auditory tasks as a Spanish
However, an informal interview with Spanish speaking subjects and in-
dependent interviews with local, bilingual speech pathologists suggests
that no significant number of test or response items were unfamiliar. The
reader is cautioned to explore such possibilities before adapting Spanish
lists to local use.
Testing materials were video-tape recordings of the MRT and SMRT. The MRT was taken from the original list of 50 six-word ensembles. These word ensembles were randomized three times resulting in three separate test lists. The SMRT was taken from the original ensembles which did not yield significant differences and compiled and randomized to produce three lists of 50 four-word ensembles. Typed response sheets were used for both tests.

The speaker for the tapes was a seasoned male newscaster from the local area and considered to be fluent in both languages for both precise phonetics and overall impression of speech. The recordings of the master tape (Scotch 3M) were made in a sound isolated TV studio with a TV camera (Norelco LDH-1) and desk microphone (Electrovoice 654) with an attached VU meter. The recorded image was a view of the head and upper shoulders against a blue background. Following the recording, each list was dubbed (Sony recorder VO-1600) onto a videocassette (Karex SCA-30).

Each test item was preceded by the carrier phrase, “Number _____; mark the Word ______;” or “Número _______; marca la palabra ______;”, with “word” and the second syllable of “palabra” peaking at OVU. The test word was delivered in a normal conversational manner. An interval of 10 seconds was allotted between each test word and the succeeding carrier phrase.

Equipment for test presentation included a videotape player (JVC Model GP-3000U); a 17-inch video screen (Sony Model 1710) and a two-channel speech audiometer (Grausen-Stadler TS). The signal was mixed at the earphone (TDH-39 mounted in an MX41/AR cushion) with a speech spectrum noise.

Each group of six subjects was divided in half and received either monaural auditory only or monaural audiovisual test presentations. The auditory only condition presented the speech signal at S/N ratios of +12, +6 and 0 dB with the speech noise held constant at 30 dB HTL. Audiovisual presentations were considered to be visual signal only, +6 and 6 dB S/N audiovisual signals. List order was counterbalanced within each group of thirty. Subjects sat in a chair three feet from a 17-inch television screen. The percentage of correctly identified items was determined for each list.

RESULTS

Limitations imposed by the number of subjects available prevented a direct examination of list effects although none were expected. An indirect examination was conducted by examining the rank order of the three lists for each test in each of the six test conditions. For the MRT, list one
produced the highest scores four times and list three produced the lowest score five times. However, taking the conservative criterion of a difference in means which exceeded only one standard deviation, list one was equal to three in three cases, was poorer than three in one case and was better than three in two cases. Considering the small sample size involved (10 per cell), it is judged that major differences among the three lists were not manifested. A similar examination of the SMRT suggested no pattern in the rank order which might lead to the conclusion that there were major list differences.

Table 1 presents the mean percentage correct and the standard deviation for the six conditions of each of the two tests, MRT and SMRT. As expected, increasingly more positive signal-to-noise ratios improved performance for purely auditory presentations. Under the parameters selected, it would appear that the upper portion of the articulation function was explored. Zero dB signal-to-noise ratios produced approximately 50% scores with smaller degrees of improvement between +6 and +12 dB than between 0 and +6 dB signal-to-noise ratios. This suggests that the articulation function had begun to plateau by +12 dB.

### Table 1. Mean Scores (%) and Standard Deviations from Three Auditory and Three Audiovisual Conditions for Both the MRT and SMRT Materials. N = 30 per cell.

<table>
<thead>
<tr>
<th></th>
<th><strong>AUDITORY</strong></th>
<th><strong>AUDIOVISUAL</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>S/N Ratio</td>
<td>0  +6  +12</td>
<td>Vision V &amp; O V &amp; +6</td>
</tr>
<tr>
<td>MRT X</td>
<td>47.4 70.9 79.7</td>
<td>55.9 77.1 87.0</td>
</tr>
<tr>
<td>SD</td>
<td>10.6 8.4 8.4</td>
<td>9.6 11.8 9.3</td>
</tr>
<tr>
<td>SMRT X</td>
<td>52.0 73.9 88.1</td>
<td>70.0 77.1 81.9</td>
</tr>
<tr>
<td>SD</td>
<td>13.2 13.6 10.0</td>
<td>13.8 13.0 10.2</td>
</tr>
</tbody>
</table>

For audiovisual presentations, the MRT Produced a function similar to that for auditory only presentations. Vision alone produced scores near 50%. The pattern of results suggested that the opportunity to lipread the materials was equivalent to an increase of more than 6 dB in signal-to-noise ratio, i.e. 0 dB audiovisual scores were higher than +6 dB auditory; only +6 dB audiovisual higher than +12 dB auditory only. It should also be noted that visual only presentations produced scores which approximated scores for normals in other tests of lipreading (Jeffers and Barley, 1977, p. 340-41; Neely, 1956). For the SMRT, less satisfactory results were obtained. The function was much shallower than hoped for
with only some 12% difference between vision only and the most favor-
able audiovisual presentation. In fact, listening alone was superior to
two of the auditory only presentations. It should be noted that the best of
the audiovisual presentations produced lower scores than the best audi-
tory only scores, an outcome unlike the results from the MRT.

Turning from mean performances in the various conditions, two over-
all trends were seen in the standard deviations. As expected, more diffi-
cult conditions produced increased variability. More importantly, for
application to clinical practice, the SMRT produced larger standard
deviations than the MRT in all six possible comparisons.

Following this general examination of results, specific hypotheses were
subjected to statistical examination:

1. Were significant improvements observed with the introduction of a
visual component to the signal? The question applied to the 0 and
+6 dB presentations, with and without visual input for both the
MRT and SMRT.

2. Was the shallower articulation function for the audiovisual pre-
sentations of the SMRT of such a degree that there were no signifi-
cant differences among the three conditions?

3. Was the increased variability of the SMRT significant?

Since answers to the above involved multiple comparisons, a Newman-
Keuls analysis for differences among means was calculated for each test.
In answer to the first question, audiovisual presentation were signifi-
cantly higher than auditory only presentations (p < .05). Only the audio-
visual presentation of the MRT at 0 dB and its auditory only +12 dB
presentation could be considered to come from the same population,
supporting the initial observation that visual cues operate, by analogy, as
if the signal-to-noise ratio had been improved by at least 6 dB. The rate of
improvement was not so clearly defined for the SMRT. In these cases any
auditory addition to vision (+6 or +12 dB) improved performance signifi-
cantly (p < .05) but the audiovisual +6 or 0 dB conditions were not
different from each other. The last answer was sought by insuring that
there was homogeneity of variance within all means using the MRT
and within all means using the SMRT. In spite of the apparent trend
toward increasing variability with increasing difficulty, Bartlett's test
judged that each set of test results was homogeneous (F(5,174 = 1.06 and
1.17, MRT and SMRT). On that basis, variance for all means with
each test was calculated and the ratio between them evaluated for
significance by using F. The outcome indicated that the two measures of
variance were not significantly different (F(1,170 = 1.62). Thus the
observation of increased variability with the SMRT was not judged to be
significant.
DISCUSSION

Four aspects of the results bear further discussion. They are:
1. the comparability of present performance measures to earlier ones,
2. the suitability of the MRT and SMRT for diagnostic purposes,
3. their suitability for rehabilitative purposes, and
4. an outline of our recommendations for their routine use.

In the absence of reports verifying initial findings with the SMRT, an unambiguous picture of its consistency among laboratories and in clinical practice cannot be had. Returning to the original report, Tosi reported a mean of 75.6% correct for a signal to noise ratio of +6 dB with white noise. This result is consistent with present findings (73.9 and 88.1% for +6 and +12 dB). Reports of lipreadability measures could not be uncovered. For the MRT, auditory only performance at +6 and +12 dB was approximately 71 and 78% which is in good agreement with measures obtained by Kreul, et al. Those investigators reported scores of 64.73 and 81-85% for similar signal-to-noise ratios. Although reports of lipreading performance with the MRT were not uncovered, earlier mentioned reports of single word tests (vs. sentences) suggest that without contextual cues, such as a multiple-guess format, performance is in the range of 25%, a score so low as to produce a "basement" rather than a "ceiling" effect. The addition of context via multiple-guess responses has produced scores comparable to those obtained here. The general improvement in performance with audiovisual presentation observed here is typical of numerous reports which need not be reviewed.

The suitability of present materials for diagnostic purposes depends upon the orientation of the clinician. If a major concern is the separation of conductive and normal hearing individuals from those who may be coarsely categorized as "sensorineural", then criteria can be leveled at the MRT and, by inference, the SMRT. In fact reports suggest that they are not as sensitive to differences among individuals with sensorineural loss as are such materials as the Bush-Hughes recordings. However, it is suggested, such distinctions are no longer of overriding importance in a diagnostic strategy which seeks to identify site of lesion and should include the wide variety of more sophisticated tests which have come to supersede simple measures of discrimination. If not in the realm of suprathreshold tests such as ipsi- and contra-lateral acoustic reflex testing and the like, modern developments in audiology have come to demonstrate that the shape of the articulation function is of greater import than any performance measure made at a single level of presentation. There is little reason to believe that the MRT is more resistant to the articulation function "roll-over" than other materials. We conclude that the two sets of test materials are satisfactory for the current status of site of lesion testing.
One of the limiting factors in the use of materials for the specific goal of gauging need for rehabilitative measures is that the bulk of such materials require additional testing beyond the presentation of “diagnostic” tests. Further, their performance with visual only presentation leaves much to be desired. The present results clearly support the proposition that the MRT and SMRT can be used for evaluation of lipreading as well as discrimination. Our findings suggest that the facilitative interaction between vision and audition is not as simple with the SMRT as with the MRT. At this point, we hesitate to suggest its unqualified use for measuring audiovisual interaction. What is obviously needed is a careful examination of the viability of the SMRT materials and their relationship to the foils which were used on the response sheets. That analysis has not begun and we offer our apologies to the reader for that lapse. We do conclude that the MRT is a satisfactory instrument for rehabilitative evaluations and, by virtue of its reported resistance to learning effects, for periodic evaluation of progress in rehabilitative processes. At least, the SMRT provides an initial approximation of a lipreading test in Spanish. Given an individual who is able to be a talker for videotaping, the test can be scored for both auditory only and visual only conditions (if not audio-visual) by the clinician unfamiliar with Spanish.

Recommendations regarding the use of the two tests fall into two categories: diagnostic and rehabilitative. For the first, we suggest that all individuals be tested at a signal-to-noise ratio of +6 dB. At this combination results are free of the ceiling effect observed when using most available materials in quiet with normals and conductive hearing losses. Deviations from normal should be readily apparent even for those individuals with mild or high frequency losses who might otherwise produce normal results. More generally, the literature is abundant with reports that discrimination materials are less efficient in quiet than in noise. The use of materials for the detection of articulation function roll-over has not been verified through, again, there is no reason to believe that they are roll-over resistant.

The advantage of the materials comes in their ability to replicate the tests under audiovisual or visual only conditions without re-examining auditory only performance. While current communication efficiency of any individual would be a function of the sensation level of the auditory materials, appropriate levels should be easily discerned from basic audio-logical information already collected. At least an audiovisual presentation at +6 dB should be made. If such a measure indicates poorer than normal performance, instruction in lipreading with concomitant use of auditory cues may be appropriate. If performance is in the range of normal, perhaps simple amplification is all that is needed. As absolute measures, the value of visual only scores is not clear. On the assumption
that normals are an unreasonable standard for lipreading proficiency since they are unlikely to depend on visual information, what can be understood from a finding that a particular patient is better or poorer than our norms? At least the visual only scores can provide an estimate of an individual's current level of performance and as a gauge of improvement during training.

CONCLUSIONS

The knowledgeable reader could have easily concluded that no report including this one, could be the final word on evaluation of communication efficiency. We believe it provides a step toward more reliable and clinically applicable techniques in addition to making relatively small demands for time in an already crowded audiological evaluation process. Equipment demands are relatively modest by brain stem audiometry standards and certainly more widely available and applicable.

1. The MRT is an appropriate tool for measuring auditory, visual and audiovisual performance.

2. The SMRT is an appropriate tool for measuring auditory performance. It is likely to be appropriate for measuring visual performance. Further exploration of its audiovisual performance and likely modification of the foils is warranted.

ACKNOWLEDGEMENT

This study was supported, in part, by the National Institutes of Health Grant # CA 18629.

REFERENCES


KOPRA, L. L., BLOSSER, D., and WALDRON, D. L. Comparison of Fairbanks Rhyme Test and CID Auditory Test W-22 in normal hear-
ing impaired listeners. J. Speech Hearing Res. 11:735-739 (1968).
KREUL, E. J., NIXON, J. C., SBYTER, K. D., BELL, D. W., and
LANG, I. S. A proposed clinical test of speech discrimination. J.
Speech Hearing Res. 11:536-552 (1968).
LOVINIC, J. H., BURGI, E. J., and CURRY, E. T. A comparative
evaluation of five speech discrimination measures. J. Speech Hearing
Res. 11:372-381 (1968).
NEELY, K. R. Effect of visual factors on the intelligibility of speech. J.
SANDERS, D. A., and GOODRICH, S. J. The relative contribution of
visual and auditory components of speech to speech intelligibility as a
function of three conditions of frequency distortion. J. Speech Hear-
TOSI, O. E. Estudio experimental sobre la inteligibilidad de un test de