Record: 1

Title: Evaluation of the speech intelligibility rating (SIR) test for hearing aid comparisons.

Authors: McDaniel, D. Michael
Cox, Robyn M.

Source: Journal of Speech & Hearing Research. Jun92, Vol. 35 Issue 3, p686. 8p. 3 Charts, 6 Graphs.

Document Type: Article

Subject Terms: *Hearing aids

NAICS/Industry Codes: 446199 All Other Health and Personal Care Stores
423450 Medical, Dental, and Hospital Equipment and Supplies Merchant Wholesalers
417930 Professional machinery, equipment and supplies merchant wholesalers
334510 Electromedical and Electrotherapeutic Apparatus Manufacturing

Abstract: Presents a study that evaluates the effectiveness of the speech intelligibility rating (SIR) test in differentiating among hearing aids. Research questions asked; Subjects; Method; Results; Discussion.

Full Text Word Count: 5985

ISSN: 0022-4685

DOI: 10.1044/jshr.3503.686

Accession Number: 9212211618

Database: Communication & Mass Media Complete

Evaluation of the Speech Intelligibility Rating (SIR) Test for Hearing Aid Comparisons

The SIR test was created for use in hearing aid comparisons. The test protocol obtains listener judgments of the intelligibility of connected speech passages. This study was conducted to evaluate the effectiveness of the SIR test in differentiating among hearing aids. Specific research questions were (a) Is the sensitivity of the SIR test sufficient for differentiating among very similar and slightly dissimilar hearing aids? (b) Does the SIR test result in reliable hearing aid rankings? and (c) What are the effects of using shortened connected speech passages? Ten listeners with hearing impairments rated the intelligibility of both full-length and shortened SIR test passages while wearing each of four individually selected hearing aids representing three different frequency/gain prescriptions. Results suggested that the SIR test is capable of differentiating among slightly dissimilar hearing aids and that hearing aid rankings resulting from speech intelligibility ratings were reliable. The decision to use full-length or shortened SIR test passages depends on the outcome the user wishes to maximize. Under the conditions used in this study, maximum sensitivity was achieved with ratings from five shortened passages, whereas maximum reliability was obtained with three full-length passages.

KEY WORDS: speech, intelligibility, ratings, hearing aid, comparisons

One goal of clinical hearing aid evaluation is to reliably stratify performance with different hearing aids. To lend validity to the evaluation process, tests for comparison of hearing aids have used speech signals for which intelligibility is scored. One such test is the Speech Intelligibility Rating (SIR) test (Cox & McDaniel, 1989). The

SIR test consists of passages of connected speech for which intelligibility is rated on an equal-appearing interval scale ranging from 0 to 10.

In a prior study, Cox and McDaniel (1984) had demonstrated that normal-hearing listeners were able to rate the intelligibility of connected speech passages using an interval scale. They found that intelligibility ratings for passages of hearing-aid-processed connected speech were able to identify two superior instruments from a pool of four similar hearing aids. However, they reported that hearing aid rankings seemed somewhat dependent on the test-signal-to-competition ratio as well as on the talker who recorded the passages. On the basis of these findings, Cox and McDaniel (1989) developed the SIR test.

The SIR test uses 20 test passages, spoken by a single talker, that were identified as having equivalent intelligibility. To facilitate use of the SIR test, a critical difference (CD), expressed in rating scale intervals, was developed. The 90% CD for normal-hearing listeners, based on the mean ratings from three passages, was 1.8 scale intervals. A 93% level of certainty was achieved by increasing the CD to 2 scale intervals.

Cox and McDaniel (1989) suggested that future study with the SIR test should be directed at providing clear guidelines for interpreting the significance of differences between hearing aid ratings for listeners with hearing impairments. Specifically, it was felt that a study was needed to (a) assess the sensitivity of the ratings to differences among aided conditions, (b) determine the number of ratings necessary to produce reliable hearing aid rankings, (c) estimate critical differences appropriate for hearing-impaired subjects, and (d) explore the use of shorter passages. These issues formed the purpose of the investigation reported here.

Method

Test Material

The test material for this investigation was the SIR test described by Cox and McDaniel (1989). As was mentioned previously, the SIR test utilizes connected speech for which intelligibility is rated on an equal-appearing interval scale. All connected speech passages are equated on the basis of subject matter, vocabulary, sentence structure, and reading level. With the exception of one 2-min "setup" passage, all passages are uniform in length (108-110 words) and require approximately 48 sec to read aloud. The topic of the passage is always mentioned within the first four words and is repeated three to seven times throughout the passage. Passages are constructed with a logical break near their middle to allow for administration of half-length passages. The SIR test is recorded on two tracks, with one track containing intensity-equalized recorded speech material and the second track consisting of a multivoice babble. A calibration noise is provided on both tracks.

The SIR test comprises four different categories of passages. One passage category is the "setup" passage. This category contains only one passage, which is approximately 2 min long. This passage is used in setting the signal-to-babble ratio for the test and for providing familiarization with the rating task. A second category of passages is identified as practice passages. Cox and McDaniel (1989) determined that these four passages, although equivalent in terms of rating means and standard deviations, were slightly more difficult than the test passages. The purpose of these passages is to substantiate the test-signal-to-babble ratio and to more thoroughly familiarize listeners with the response task. Another category of passages is the 20 test passages. These passages are used to obtain intelligibility ratings for different hearing aids or hearing aid conditions. The score assigned to a hearing aid or hearing aid condition is the average rating of three to five test passages. The fourth category of passages contains five passages, considered to be validity passages. This group of passages has been shown to elicit a wide range of mean ratings from normal-hearing listeners (Cox & McDaniel, 1989). Within this category is a "catch" passage that is recorded backwards and therefore is totally

unintelligible. These validity passages are used to verify that listeners are responding on the basis of intelligibility. Listeners who do not respond to these passages in the appropriate order, or nearly so, are not good candidates for the SIR test.

Test passages are presented at an individually established signal-to-babble ratio. The SIR test protocol calls for a test-signal-to-babble ratio that will result in intelligibility ratings between 7 and 8 on a 10-point scale for portions of the setup passage. This test-signal-to-babble ratio is established by a bracketing technique described in an earlier article (see Cox & McDaniel, 1989, Figure 4). Briefly, this procedure is designed to efficiently delineate the test-signal-to-babble ratio while maintaining the overall presentation level. Subjects are asked to rate the intelligibility of a 20-see portion of the setup passage. Depending on the outcome of that rating, the competing babble is adjusted in 5-dB steps to a level for which the rating is closer to 7 or 8. Subsequent ratings are obtained at a smaller (2.5-dB) step size until a signal-tobabble ratio is achieved that results in ratings of 7 or 8 for 20-see portions of the setup passage.

Subjects

Subjects for this study were 5 female and 5 male hearing-impaired adults aged 22-74 years (mean age = 61.7 years). Each subject had a sensorineural hearing impairment with a three-frequency pure-tone average (500,1000, and 2000 Hz) for the test ear no better than 30 dB HL (ANSI, 1969). If a subject's hearing impairment was not bilaterally symmetrical, the better ear was used for testing. Table 1 summarizes the audiometric data for each of the test ears used in this study.

Hearing Aids

Three different frequency/gain prescriptions were developed for each subject. The first of these prescriptions (M) was derived directly from a hearing aid prescription procedure (MSU Version 2, Cox, 1985). The second and third prescriptions were generated by modifying the slope of the MSU prescription. For the second prescription (H), the MSU prescription was modified to reduce the low-frequency gain. This modification resulted in a +3 dB/octave steeper slope relative to the MSU prescription. The third prescription (L) represented a modification of the MSU prescription resulting in less high-frequency gain. This third prescription produced a -3 dB/octave less steep slope relative to the MSU prescription.

Four hearing aid conditions were selected. Two of these conditions, arbitrarily assigned the labels of M1 and M2, matched each subject's MSU prescription. The remaining two hearing aid conditions conformed to the modified prescriptions labelled L and H. Four separate hearing aids, representing the four conditions, were selected for each subject for use in this study. A total of 19 different hearing aids representing 14 models supplied by five different manufacturers were used. Before using the hearing aids in this study, we determined that they were all functioning in accordance with the manufacturers' specifications.

All hearing aids were behind-the-ear models equipped with directional microphones and linear amplifiers. The saturation output was set at maximum for each hearing aid. We chose this maximum output selection because the subjects were in a controlled acoustic environment and we felt they would not be exposed to uncomfortably loud amplified sounds. The typical high-frequency average saturation sound pressure level for a 90-dB input for the instruments used in this study was greater than 120 dB (range = 108-130 dB). The adjusted maximum gain at any test frequency for any instrument used in this study did not exceed 35 dB. By using the maximum output selection for each instrument, we felt that any compromise to intelligibility due to peak clipping would be minimized or avoided.

Given the constraints imposed by the hearing aids, it was not possible to obtain instruments that exactly conformed to the prescriptions. A hearing aid was judged to conform to the prescription if its actual root mean

square (RMS) deviation for seven test frequencies (250, 500, 800, 1000, 1600, 2500, and 4000 Hz), as measured in a 2-cc coupler, was plus or minus 3 dB of the prescription. We chose the + 3-dB tolerance value on the basis of data presented by Dillon (1985), who reported that when a frequency response differed from an optimal response by an RMS deviation of 3 dB, about one half of his subjects judged that response to be inferior in terms of speech intelligibility. We recognized that if we used coupler gain, the actual insertion gain for each instrument was not known. However, because all subjects used unvented earmolds, the slope relationships among the four conditions would be the same in both coupler and real-ear measurements.

To achieve the desired coupler gain for the test hearing aid conditions, instruments with a wide range of low-cut tone controls were used. Several custom earmolds were made for each subject, with a variety of horn effects and damper configurations for use as needed to achieve the gain requirements. All hearing aid measurements were made with their earmolds situated in an HA 1 coupler. All user and fitter controls, including volume control, were secured for hearing aids M1 and M2 after the instruments were adjusted to produce the requirements specified by the prescription. Hearing aids H and L were selected so that, following adjustments to match the prescriptions, there was at least 10 dB of reserve gain at all frequencies. For these hearing aids, the volume controls were unsecured to allow for later adjustment.

Instrumentation

Test passages were replayed by an optical disk player (Panasonic model TQ 2024F) that was controlled by a microcomputer (Apple lie). The Channel 1 and Channel 2 audio outputs were delivered to the respective inputs of a two-channel clinical audiometer. Outputs of the audiometer were routed through a custom mixer, an active equalizer (White, series 4000 model 4002) adjusted to compensate for loudspeaker frequency response irregularities, and a power amplifier (Macintosh model 250). Test passages and babble were delivered to a single loudspeaker (Realistic Minimus 7) located in a double-wall sound room. The frequency response of the entire signal delivery system was flat (+/- 5 dB) from 200 through 10,000 Hz when measured in the sound field at the subject position.

Procedure

Subjects were seated in the sound field, at a 0 Degrees azimuth, 1 m from the loudspeaker. The nontest ear was plugged and muffed.

For each subject, it was necessary to equate the hearing aid conditions for loudness. Hearing aids M1 and M2 were already presumed to be at an optimum loudness level. However, instruments L and H, which represented modified prescriptions, were likely set during the prescription-matching process to provide a level lower than optimal. To equate the loudness of amplified speech for all four hearing aids a protocol was employed in which subjects compared the loudness of a reference speech spectrum noise at 55 dB Leq (approximately 30 see) heard through hearing aid M1, with the same noise amplified by hearing aids L and H.

As part of this loudness-matching protocol, subjects adjusted the volume controls of hearing aids L and H to produce a loudness equal to hearing aid M1. A pilot study with this type of procedure had shown that hearing-impaired subjects were able to perform the task with considerable reliability. Four adjustments were obtained for each instrument using a bracketing procedure. Two adjustments were made from a volume control setting selected to cause a higher loudness level than that of the reference noise and two from a volume control setting selected to result in a lower loudness level. After each loudness-matching adjustment, the instrument was removed and the gain at 1000 Hz was recorded. These four gain values were averaged, and the instrument's volume control was set to a position that resulted in the average gain value. The average standard deviation of these loudness-matching adjustments, expressed in dB gain at 1000 Hz, was 2.34, with a range of

1.09-3.54 dB. Following the loudness-matching adjustments, the volume controls for hearing aids L and H were secured.

Figure 1 shows the mean 2-cc coupler gain values for each of the four hearing aid conditions used in this study. The mean gain-by-frequency prescription values indicated by MSU prescription Version 2 are also given in Figure 1.

Subjects were given written instructions for the SIR test. They were instructed to listen to each passage and then assign that passage a rating corresponding to the overall intelligibility of the passage on the equal-appearing interval scale. Figure 2 shows the equal-appearing interval scale as it was used in this study. Instructions were the same as those used by Cox and McDaniel (1989) except that subjects were allowed to mark between the numbers. The test passages were presented at a level of 55 dB Leq (approximately 30 see).

During the loudness-matching procedure, the establishment of the test-signal-to-babble ratio, and actual testing, we attempted to reduce or prevent any bias for a particular instrument by not permitting subjects to see the hearing aids they were wearing. The M1 hearing aid was worn to establish the test-signal-to-babble ratio and for the practice and validity passage presentation. Data were collected for each subject on 2 test days separated by a minimum of 13 days. Half of the subjects listened to the full-length test passages on the first day and half listened to the half-length test passages on the first day. The order was reversed for the second test day. On the second test day a portion of the setup passage, in conjunction with the practice passages, was used to reorient the subjects to the task. Subjects wore hearing aid M1 for this reorientation. For each of the two test days a total of five test passages was administered for each hearing aid condition. Thus, 20 test intelligibility ratings were obtained for each test day. Hearing aids were removed after each rating. The order of hearing aid presentation was counterbalanced. For each subject on each test day, the 20 test passages were presented in a random order without replacement.

Results

Before analyzing data to answer specific research questions, we were interested in knowing whether any variable other than frequency responses influenced the final rankings of the four hearing aids used for each subject. Two factors that could have created a bias were a preference for some particular hearing aid model or a preference for some particular earmold damping configuration. Tabulation of the final hearing aid rankings across subjects by hearing aid model and damping configuration suggested no discernible preference that might have influenced our subjects' intelligibility ratings.

Though not the same for each subject, each of the hearing aid prescriptions was derived under the same set of conditions; thus, we refer to hearing aid conditions. Figure 3 is the composite mean rating score for both full- and half-length passages for each of the four hearing aid conditions. To determine the overall sensitivity of the SIR test to similar as well as slightly dissimilar hearing aid conditions, an analysis of variance (ANOVA) for repeated measures was performed on the mean rating score for the five passage ratings for each hearing aid condition. Variables for the ANOVA were passage length (full- vs. half-length), hearing aid conditions (L, M1, M2, and H), and subjects. Results of the ANOVA indicated that only the main effect of hearing aid conditions was significant, $F(3, 27) = 32.371$, $p < .00001$.

A number of comparisons were made to address specific research questions: hearing aid condition M1 versus hearing aid condition M2 (similar frequency response); hearing aid condition M1 + M2 versus hearing aid condition H (3 dB/octave steeper slope); hearing aid condition M1 + M2 versus hearing aid condition L (3 dB/octave less steep slope); and finally, hearing aid condition H versus hearing aid condition L (6 dB/octave difference). A Scheffe test was used for these comparisons (Scheffe, 1953). Results presented in Table 2 show

essentially the same findings for both passage lengths and indicate that the ratings for the M hearing aid conditions were not significantly different from one another. Likewise, hearing aid conditions H and L produced intelligibility ratings that were not significantly different from one another. However, the compound contrast that pitted the M hearing aid conditions against hearing aid condition L indicated that the intelligibility ratings from these conditions were significantly different for both passage lengths. Likewise, the compound contrast that pitted the M hearing aid conditions against hearing aid condition H indicated that the intelligibility ratings for these conditions were significantly different.

To explore the sensitivity of the SIR test further in a manner that could be easily adapted for clinical use, 90% critical differences (CDs) were developed. A 90% confidence criterion was selected because the SIR test was designed to be a clinical tool and the selection of a best instrument 9 out of 10 times seemed an enviable goal. CDs were developed for both passage lengths based on five ratings per hearing aid condition. A variation of the Fisher least significant difference (LSD) formula (Fisher, 1937) was used to establish the CDs. This procedure required a determination of the typical within-subject variability for hearing aid intelligibility ratings. To accomplish this, a single-factor ANOVA was performed on the four hearing aid conditions for each subject for each passage length. Typical error terms for each passage length were derived by averaging the 10 individual error terms. These error terms were then entered into the Fisher equation to derive the LSD between two mean ratings for a 90% level of certainty. The obtained CDs were 1.75 scale intervals for the full-length passages and 1.98 scale intervals for the half-length passages.

Having derived CDs for the SIR test, we wished to evaluate the outcome of the present investigation when the CDs were applied to the individual subject's ratings. Specifically, we determined for each subject whether the hearing aid conditions that were ranked second and third were rated lower than the highest ranked hearing aid condition by an amount that exceeded the CD. For this evaluation, ratings for the two M hearing aid conditions were averaged. Figure 4 presents the proportion of subjects for which mean ratings for the various hearing aid conditions differed by more than the 90% CD.

Results presented in Figure 4 suggest that the SIR test, using a 90% CD, was completely effective in differentiating between the hearing aid condition resulting in the first (highest) ranking and the hearing aid condition that was ranked third. These two conditions yielded significantly different mean ratings for all 10 subjects for both the full- and half-length passages. However, the differentiation between the first- and second-ranked hearing aid condition was not as clear. Results indicated that the best differentiation between these two conditions was obtained with the half-length passages.

We used a correlation approach to determine the number of intelligibility ratings required per hearing aid condition to achieve reliable rankings for each passage length. The four hearing aid conditions were ranked after each passage presentation using their cumulative rating score. Rankings were derived for each subject. The final ranking, which was based on five passages, was taken to be the best estimate of the true ranking for the four hearing aid conditions. Spearman correlation coefficients were computed between the final ranking and each of the first four cumulative rankings.

To illustrate these findings, a cumulative distribution was constructed. This distribution reflects the number of passage intelligibility ratings required to achieve a correlation of .95 or better with the estimated true ranking of the four hearing aid conditions. Figure 5 shows the cumulative distributions for both the full- and half-length passages. As Figure 5 shows, for the full-length passages, all 10 subjects achieved the criterion of .95 by the third passage. For the half-length passages, 7 of the 10 subjects reached the criterion correlation by the third passage. It should be noted that these correlations included both exemplars of hearing aid condition M. For 2

of the 3 subjects who failed to reach a .95 correlation by the third half-length passage, the only change was an interchange of positions between hearing aid conditions M1 and M2. Nevertheless, these findings seem to suggest that ratings for the half-length passages do not possess quite as much reliability as ratings for the full-length passages.

As a second means of exploring the reliability of hearing aid rankings from intelligibility ratings, we examined the test-retest reliability of the rankings. Because the group ANOVA resulted in no significant main effect for passage length [$F(1, 9) = .024$, $p > .0000$], data for the two passage length sessions were treated as if they were simply a retest of the same condition. Final hearing aid condition rankings were averaged across subjects for both passage lengths. Mean rankings for the four hearing aid conditions are shown in Figure 6. The only rank difference for the hearing aid conditions was a change in the first-place position for the two M hearing aid conditions. For the full-length passages, hearing aid condition M2 was ranked first, and hearing aid condition M1 was ranked first for the half-length passages. A Spearman rank order correlation coefficient computed between the two ranks was .97. If the two M hearing aid conditions were treated as a single condition, the overall correlation for the two test sessions would be higher.

One related issue pertaining to the SIR test involves responses to the validity passages, which were selected because they resulted in intelligibility ratings from normal hearing listeners and spanned the range of difficulty represented by the equal-appearing interval scale. If listeners do not produce intelligibility ratings for these validity passages that approximate the same overall order of difficulty, then some question is raised as to whether or not the listeners were basing their ratings on perceived intelligibility. No subjects were excluded from this study because of their inability to respond appropriately on the basis of perceived intelligibility. Table 3 presents the group results of intelligibility ratings for these validity passages obtained with hearing aid condition M1. For comparison, results from four of these same validity passages are presented for normal-hearing listeners from Cox and McDaniel (1989). Data for Passage 1, which is the "catch" passage, were not available for the normal-hearing listeners. Although results presented in Table 3 do not indicate exactly the same average rating score for these passages for normal-hearing and for hearing-impaired listeners, the overall positioning of the ratings is the same. This finding supports the assumption that the hearing-impaired listeners in this investigation were responding on the basis of perceived intelligibility.

Discussion

In a review of the applicable evidence involving comparisons of hearing aids with similar frequency/gain slopes, Schwartz and Walden (1983) stated, "As long as the patient is being evaluated with a series of hearing aids with similar electroacoustic response characteristics, there is little reason to believe that significant differences among these homogeneous instruments should emerge on any clinical measure" (p. 348). Results for hearing aid conditions M1 and M2 from this study appear to be consistent with this statement by Schwartz and Walden. There is still no reason to believe that significant speech intelligibility differences will emerge between well-functioning hearing aids with highly similar frequency/gain characteristics. However, in spite of this outcome, there may still be valid reasons for comparing hearing aids that conform to the same prescription. Clinical hearing aid comparisons may still be the most practical and efficient method for assessing speech intelligibility with different signal processing methods such as syllabic compression and automatic low-frequency reduction in noise.

Perhaps the most practical application for the SIR test is to evaluate hearing aids or hearing aid conditions that are dissimilar. Changes in prescribed frequency responses, which are becoming increasingly possible with programmable hearing aids, may substantially alter perceived intelligibility to the listeners. Thus, to maximize aided performance, the need still exists to measure speech intelligibility through two or more frequency

responses. Results from this evaluation of the SIR test indicated the test was able to differentiate between the M hearing aid conditions and those with a relative +3 dB/octave slope and a relative -3 dB/octave slope. These findings suggest the SIR test may be of benefit in evaluating hearing aid or hearing aid conditions that exhibit slightly different frequency/gain characteristics.

Results from this investigation have an apparent contradiction. Hearing aid conditions that differed by a slope difference of 3 dB/octave were found to be different, yet hearing aid conditions L and H, which differed by a slope difference of 6 dB/octave, were found to be no different. A review of the rankings for five passages revealed there was no clear trend for either hearing aid condition L or H to be ranked higher. This finding indicates that, relative to hearing aid condition M, hearing aid conditions H and L were equally poor and thus no differences emerged from intelligibility ratings for the SIR test.

The outcome from the present study underscores the importance of selecting initial hearing aid test conditions for this type of study. It conditions chosen for the present study had not incorporated the M frequency response but instead had used the L, H, and other frequency responses with more widely divergent slopes, the results probably would have suggested that the SIR test was rather insensitive, being unable to differentiate between two instruments that differed by 6 dB/octave.

Another way to view the sensitivity of a speech intelligibility test in hearing aid comparisons is through the reliability of its hearing aid rankings. Individual results from this investigation suggested that the reliability of rankings for hearing aid conditions was, to some extent, dependent on the length of the passages used. Reliable rankings of hearing aid conditions were obtained for the full-length passages with three intelligibility ratings per hearing aid condition. However, when the half-length passages were used, more than four ratings were required for equal reliability (see Figure 5). These seemingly different findings were, to some extent, a product of the fact that hearing aid conditions M1 and M2 were similar. Their similarity resulted in occasional transpositions of their rankings, which must be interpreted as chance since they had already been shown to be no different from each other.

The implication of this finding is that the SIR test appears to result in relatively reliable hearing aid rankings after three full-length passages. Three ratings can be obtained in a period of time that is well suited to a conventional hearing aid evaluation without placing undue hardship on older listeners who have hearing impairments.

The 90% CD reported by Cox and McDaniel (1989) for normal-hearing listeners, which was based on the average of three intelligibility ratings, was 1.8 scale intervals. The 90% CDs from this study, based on five ratings for both the full-and half-length passages (1.75 vs. 1.98), were in close agreement with those of Cox and McDaniel. Though the size of the CD was larger for the half-length passages than for the full-length passages, the difference does not appear to be a deterrent for adopting a CD for clinical use. The adoption of a CD of 2 scale intervals seems to be a useful clinical criterion that would produce a greater than 90% confidence that two conditions are indeed different.

One of the major aims of this investigation was to try to determine whether shorter versions of the continuous discourse passages could be substituted for the full-length passages without compromising the sensitivity and reliability of the SIR test. The measure of choice is the one that will result in the greatest sensitivity and reliability in the least amount of time. The findings of this investigation indicate that the choice of passage length depends on the outcome the user wishes to maximize. If sensitivity to different hearing aid characteristics is of the utmost concern, then perhaps five shorter passages should be used for each condition. If, however, reliable hearing aid rankings from the fewest number of passage presentations is the primary

concern, then perhaps three full-length passages should be used. In either case, the time requirement of 144 sec (3 x 48) for three full-length passages and 120 sec (5 x 24) does not appear to be a major clinical concern.

Several features of the SIR test were confirmed or reaffirmed during the course of this investigation. Perhaps of most importance was the finding that none of the hearing-impaired subjects in this study experienced difficulty in understanding the instructions or the response task of the SIR test. After brief exposure to the practice passages, subjects became proficient with the task. Results from the validity passages support the contention that intelligibility ratings are a valid means of quantifying perceived speech intelligibility. It is anticipated that implementation of the SIR test in a clinical situation where three conditions were to be compared would take less than 30 min, using five full-length passages per condition. Finally, the SIR test can be performed with equipment already present in most audiological centers. All that is required is a taped, or otherwise recorded, copy of the SIR test and an audiometer with two channels that can be independently adjusted to deliver their outputs to one loudspeaker situated in a calibrated sound field.

It is well known that rapid learning occurs for contextual speech material such as that used in the SIR test passages. In addition, a significant "procedure learning" effect is associated with all speech intelligibility tests (Theodoridis & Schoeny, 1990). If valid results are to be obtained, it is important to control these effects when using the SIR test to compare hearing aids or hearing aid conditions. Precautions should be taken to minimize the effects of procedure learning by providing a sufficient amount of practice prior to testing and by distributing the remaining learning effects equally across conditions by changing conditions after administration of each test passage. Ideally, test passages should not be repeated. If repetition is unavoidable, the effects of subjects' memory of passage content should be distributed across conditions by using an equal number of repeated passages in each condition.

Conclusion

The trend of current clinical practice has been shifting away from comparative hearing aid testing. This trend seems to be a result of several factors, including the development of formula-based amplification prescriptions, the increasing use of custom in-the-ear and in-the-canal instruments, and, perhaps in part, the lack of a clinical tool that is effective in differentiating between hearing aids or hearing aid conditions. However, the need to recognize atypical gain, frequency response, and maximum output requirements as well as to identify optimal signal processing schemes requires a tool for hearing aid comparisons that is valid, reliable, and easy to administer. The findings of this investigation suggest that the SIR test may prove useful in these applications.

Acknowledgments

This work was supported in part by the Memphis State University Center for Research Initiatives and Strategies for the Communicatively Impaired (CRISCI) and by VA Rehabilitation Research and Development funds. The authors thank Genevieve Alexander for her technical support, Robert M. Joyce for the development of the software for presentation of the SIR test passages, and Ginny Crihfield-Campbell for her assistance in preparing this manuscript.

TABLE 1. Audiometric threshold (dB/HL) data for subjects.

	Frequency (Hz)					
Subject	250	500	1000	2000	4000	8000
1	15	15	50	65	60	50
2	20	35	45	50	60	60
3	45	50	60	65	80	80
4	45	40	40	55	75	85
5	25	25	45	55	95	110

6	25	35	45	55	70	75
7	45	45	50	60	50	75
8	25	25	20	45	50	60
9	30	25	35	50	55	85
10	35	35	40	50	45	55
M	31	33	43	55	64	74
SD	10.44	10.05	10.05	6.32	14.97	17.04

TABLE 2. Results for the group Scheffe contrast for both full- and half-length passages. Table values are F values (df = 3,36).

	Passage length	
Contrast	Full	Half
M1 vs. M2	0.021	0.001
M vs. L	9.697[*]	14.564[*]
M vs. H	11.650[*]	23.083[*]
H vs. L	0.067	1.731

[*] p < .001.

TABLE 3. Responses to validity passages from this study and from equivalency study (Cox & McDaniel, 1989).

Passage	1989 study	Present study
Wheat[a]	N/A	0.35
Vegetable	3.25	5.50
Lettuce	4.50	5.60
Carrot	7.20	7.10
Watch	8.00	7.65

[a] Denotes "catch" passage.

GRAPH: FIGURE 1. Mean 2-cc coupler gain values by frequency for hearing aid conditions M1, M2, H, and L. Gain values for hearing aid conditions L and H were obtained after the loudness-matching procedure. The solid line represents the mean gain-by-frequency prescription (MSU Version 2) for the 10 subjects used in this investigation.

GRAPH: FIGURE 2. Facsimile of scale used by subjects for producing Intelligibility ratings from the SIR test.

GRAPH: FIGURE 3. Mean SIR teat scores for each of the tour hearing aid conditions for both full- and halt-length passages.

GRAPH: FIGURE 4. Proportion of subjects for which mean rankings derived from the ratings of the various hearing aid conditions exceeded the 90% CD for both the full- and half-length passages. The 90% CD was 1.75 scale Intervals for the full-length passages and 1.98 scale Intervals for the halt-length passages.

GRAPH: FIGURE 5. Cumulative distributions showing the number of Intelligibility ratings required to reach the reliability criterion of .95 or better between the successive rankings and the estimated true ranking.

GRAPH: FIGURE 6. Mean rankings for the four hearing aid conditions for both passage lengths.

References

American National Standards Institute. (1969). American National Standard specification for audiometers (ANSI S3.6-1969). New York: ANSI.

Cox, R. M. (1985). ULCL-based prescriptions for in-the-ear hearing aids. Hearing Instruments, 36(4), 12-14.

Cox, R. M., & McDaniel, D. M. (1984). Intelligibility ratings of continuous discourse: Application to hearing aid selection. Journal of the Acoustical Society of America, 76, 758-766.

Cox, R. M., & McDaniel, D. M. (1989). Development of the Speech Intelligibility Rating (SIR) test for hearing aid comparisons. Journal of Speech and Hearing Research, 32, 347-352.

Dillon, H. (1985). How close is close enough? Paper presented at the National Acoustics Laboratory Hearing Aid Conference, Sydney, Australia.

Fisher, R. A. (1937). The design of experiments (2nd ed.). London: Oliver and Boyd.

Scheffe, H. (1953). A method of judging all possible contrasts in the analysis of variance. Biometrika, 40, 87-104.

Schwartz, D. M., & Walden, B. E. (1983). Speech audiometry and hearing aid assessment: A reappraisal of an old philosophy. In D. K. Konkle & W. F. Rintelmann (Eds.), Principals of speech audiometry (pp. 321-352). Baltimore: University Park Press.

Theodoridis, G. C., & Schoeny, Z. G. (1990). Procedure learning effects in speech perception tests. Audiology, 29, 228-239.

Received January 2, 1991 Accepted August 26, 1991

Contact author: D. Mike McDaniel, PhD, Department of Special Education and Communicative Disorders, Arkansas State University, P.O. Box 907, State University, AR 72467.

~~~~~~~~

By D. Michael McDaniel, Department of Special Education and Communicative Disorders, Arkansas State University, AR and Robyn M. Cox, Department of Audiology and Speech Pathology, Memphis State University

---
~~~~~~~~