Comparison of Frequency Response and Aided Speech-Recognition Performance for Hearing Aids Selected by Three Different - Prescriptive Methods

Larry Humes
Troy Hackett

Abstract

The gain and the speech-recognition performance obtained for behind-the-ear (BTE) hearing aids selected by different prescriptive methods was compared. The three prescriptive methods examined were the NAL-revised method, the Memphis State or COX method, and the POGO-II procedure. Twelve subjects had hearing aids and earmolds selected to match the prescription made by each method. Once selected, the instruments were fine tuned to match the obtained gain to that prescribed and then speech recognition was assessed. Subjects listened to the CUNY Nonsense Syllable Test (NST) in quiet and in noise. Results indicated that, although the prescribed frequency responses differed significantly among methods, there were no differences among the three prescriptive methods in obtained frequency responses or aided speech-recognition performance.

Key Words: Presbyacusis, hearing disorders, central auditory disorder, aging

For several decades, the comparative approach to hearing aid selection and evaluation, described originally by Carhart (1946), was the selection method preferred by audiologists (Burney, 1972; Smaldino and Hoene, 1981a, 1981b). Although problems inherent with this approach to hearing aid selection had been reported in the 1960s (Resnick and Becker, 1963; Shore et al, 1960), its most serious challenge came in the early 1980s. Walden et al (1983) systematically evaluated five assumptions central to the comparative approach. They found all five assumptions to be invalid for conditions representative of those under which the method was being used by audiologists.

During the past decade, there has been renewed interest in the development of prescriptive hearing aid selection methods as alternatives to the comparative procedure. At least a dozen different methods have been described since the mid 1970s. Humes (1986) evaluated 10 of these methods and found: (a) the gain prescribed was very similar for most methods, but different enough to result in the selection of different hearing aids from among several hundred responses in a large data base and (b) there was little or no difference in expected speech-recognition performance for the various methods, especially when allowing for adjustment of the volume control. Speech-recognition performance, however, was not actually measured in the study by Humes (1986). Rather, the Articulation Index (AI), an acoustical index monotonically related to speech recognition, was used to perform a theoretical evaluation of the prescribed gain characteristics. The AI, however, is not universally accepted as an accurate predictor of speech-recognition performance in the hearing impaired (Byrne, 1988).

More recently, Sullivan et al (1988) obtained actual speech-recognition scores from subjects listening to speech that had been shaped by one

Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana
Reprint requests: Larry Humes, Department of Speech and Hearing Sciences, Indiana University, Bloomington, IN 47405

of four different frequency responses. The frequency responses did not adhere strictly to any particular prescriptive method. Rather, each response was based on a principle underlying one or more of the existing selection methods. Little or no difference was observed in speech-recognition performance for speech processed by the various frequency responses, especially when the level of presentation corresponded to each listener's most comfortable loudness level. The actual frequency responses in this condition were also remarkably similar.

In the study by Sullivan et al (1988), however, real hearing aids were not used. Rather, a high-fidelity test system was used in which the desired frequency response was realized through adjustment of a digital filter and stimuli were presented over headphones. In addition, actual hearing aid prescription methods were not used; only general approximations to the rules used by some of the methods.

The question still remains, therefore, whether one existing prescriptive method is better than another, in terms of aided speech recognition, when evaluated under representative clinical conditions and with actual hearing aids. In the present study, the speech-recognition performance obtained with hearing aids selected by three contemporary prescriptive methods was evaluated. The three selection methods chosen were the NAL-R method (Byrne and Dillon, 1986), the Memphis State or COX procedure (Cox, 1988), and the POGO-II method (Schwartz et al, 1988). These methods were selected because of their common usage by audiologists, their detailed and documented validation, and the availability of computer software that incorporates all three methods (Humes, 1988). In addition to examination of between-method differences in aided speech recognition, we examined differences in prescribed and obtained frequency responses for the three selection methods.

METHOD

Subjects

Twelve listeners with sensorineural hearing loss ranging in age from 33 years to 67 years (M = 58 years) served as subjects. Pure-tone hearing thresholds for the 12 subjects were measured using an ascending modified method of limits with 4-dB descending steps and 2-dB

ascending steps. Air-conduction thresholds for each subject appear in Figure 1. Air-bone gaps were less than 10 dB from 250 through 4000 Hz and immittance measurements were normal for all subjects. The subjects have been separated into three subgroups in this figure according to their audiometric configuration. Subgroups were determined by calculating the difference between the mean low-frequency threshold (250 and 500 Hz) and the mean high-frequency threshold (4000 and 8000 Hz) for each subject. For the flat subgroup, these differences were 20 dB or less. For the steeply sloping subgroup, the
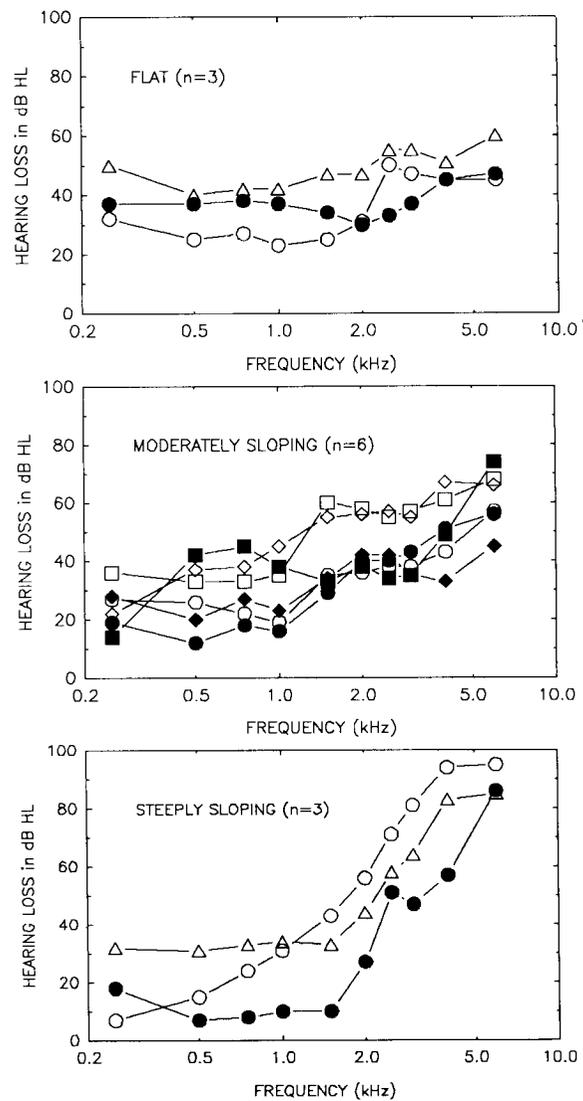

Figure 1 Hearing thresholds of the 12 listeners in this study. The subjects have been classified into three subgroups according to the configuration of their hearing loss.

differences were greater than 50 dB. The moderately sloping subgroup was comprised of all subjects not falling into either of the other two subgroups. It is clear from Figure 1 that subjects having a variety of audiograms from flat to steeply sloping participated in this study. This can be an important consideration when comparing prescribed and obtained frequency responses among various prescriptive procedures.

Procedure

Threshold and the upper limit of comfortable loudness (ULCL; Cox, 1983) were first measured under headphones for each subject. Whereas the NAL-R and POGO-II methods require only thresholds for the generation of a gain prescription, the COX method requires both threshold and ULCL. Threshold was measured at 250, 500, 750, 1000, 1500, 2000, 2500, 3000, 4000, and 6000 Hz. The ULCL was measured at all of these same frequencies, except 2000 and 3000 Hz.

SHAPE hearing aid selection software (Humes, 1988) was used to generate the prescribed real-ear gain for each of the three methods. SHAPE also selects the behind-the-ear (BTE) instrument and earmold configuration from a custom data base that yields the closest match to the prescribed gain. No prescriptions were generated for SSPL90. SSPL90 was adjusted to the minimum setting possible which would not have any affect on the gain measured for input levels from 60 to 85 dB SPL. The recommended BTE aid was always one available in our clinic stock (a data base comprised of over 300 responses). An earmold impression was made and the earmold configuration selected by the audiologist (and specified in the SHAPE software) was ordered. BTE instruments, although of decreasing popularity, were used in this study for two reasons. First, the use of BTE aids allowed us to make use of our clinic stock of instruments and incur only the cost of making three earmolds for each listener. It was not feasible to order three in-the-ear instruments for each subject. Second, the BTE aids and accompanying earmolds are more flexible electroacoustically so that the instrument could be fine tuned more easily to match the obtained real-ear gain as closely as possible to that prescribed by each selection method.

When the earmolds ordered for each instrument were received from the manufacturer, the subject was contacted and scheduled for three additional sessions, each on a separate day. During the first session, unaided speech-recognition performance was measured using the CUNY Nonsense Syllable Test (NST; Resnick et al, 1975). The 7-subtest version of this test, which consists of 62 consonant-vowel or vowel-consonant nonsense syllables administered with a closed-set response format, was used. The reliability of the NST has been established previously (Dubno and Dirks, 1982; Humes et al, 1987). The taped NST materials were presented through a clinical audiometer in the sound field at a level of 70 dB SPL from a loudspeaker positioned at 0° azimuth. For testing in a noise background, the cafeteria noise supplied with the NST was used. It was presented from the same loudspeaker at an overall sound-field level of 65 dB SPL to produce a +5 dB signal-to-noise ratio.

Following completion of the unaided speech-recognition measurements, the hearing aid and earmold selected by one of the three hearing-aid selection methods was placed on the subject. The nontest ear was occluded with a deeply seated E-A-R foam earplug. The real-ear insertion gain of the hearing aid was then measured using an Acoustimed HA–2000 test system. A click stimulus having an overall RMS sound pressure level of 70 dB was used to measure the insertion gain from 250 through 5000 Hz. The hearing aid and earmold were then modified as needed to obtain a close match to the real-ear gain prescribed by the selection method. Based on previous data on the test-retest reliability of insertion-gain measures (reviewed in Humes et al, 1988), the following goals for matching obtained gain to prescribed gain were established: (a) for frequencies less than 2000 Hz, obtained gain within 6 dB of prescribed gain; (b) for frequencies from 2000 through 3000 Hz, obtained gain within 8 dB of prescribed gain; and (c) for frequencies greater than 3000 Hz, obtained gain within 10 dB of prescribed gain. Prescribed gain at 6000 Hz available in two of the three methods (NAL-R, COX), was not considered in this study because of the inability to confirm the amount of real-ear gain at that frequency reliably. Once these fitting goals were achieved, testing proceeded to the measurement of aided speech-recognition performance.

Aided speech-recognition measures consisted of a replication of the unaided measurements. The controls of the hearing aid, including the volume control, were left unchanged from the

final settings achieved during the gain fine-tuning process.

In subsequent sessions, the gain fine-tuning process and the measurement of aided speech-recognition performance were repeated for the remaining two selection methods; a separate session for each method. The order in which each selection method was evaluated was counterbalanced across subjects.

RESULTS AND DISCUSSION

The mean prescribed and obtained gain values are plotted as a function of frequency in the top two panels of Figure 2. The COX method includes 2500 Hz as a test frequency, whereas the other two methods include 2000 and 3000 Hz as test frequencies. For the latter two methods, the gain prescribed and obtained at 2500 Hz was estimated by averaging the gain at 2000 and 3000 Hz. It is apparent from the data in the top two panels of Figure 2 that there is little difference among prescription methods in either the prescribed or the obtained insertion gain. In either panel, the difference in mean real-ear gain across methods never exceeds 10 dB at any frequency and typically approximates 5 dB. Comparing the data in the top panel to those in the middle panel, however, suggests that the obtained gain (middle panel) is consistently lower than the prescribed gain (top panel) at the highest frequency (4000 Hz).

Separate repeated-measures analyses of variance were performed on the prescribed and obtained real-ear gain values appearing in the top two panels of Figure 2. For the prescribed gain, the analysis revealed a significant effect of frequency (F (6, 66) = 7.47, p < 0.01) and selection method (F (2, 22) = 39.2, p < 0.01) and a significant interaction between these two factors (F (12, 132) = 19.4, p < 0.01). The effect of frequency on prescribed gain was expected given the sloping audiometric configurations of most subjects. The significant effects of selection method and the significant interaction between method and frequency indicate that both the prescribed absolute and the prescribed relative insertion gain (i.e., the frequency response) were not the same for all three methods.

For the obtained gain, the analysis revealed a significant main effect of frequency (F (6, 66) = 15.25, p < 0.01) and selection method (F (2, 22) = 6.0, p < 0.01) with no interaction between these two factors (F (12, 132) = 1.72, p > 0.05).
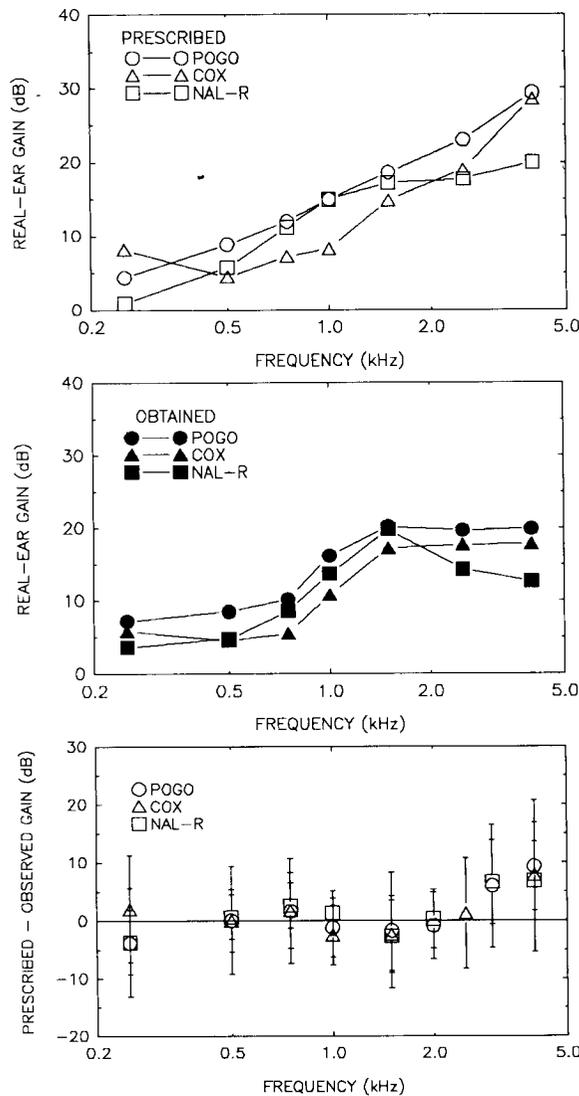

Figure 2 *Top*: mean prescribed insertion gain for each of three selection methods. *Middle*: mean obtained insertion gain for each of three selection methods. *Bottom*: means and standard deviations (*vertical bars*) of the differences between the gain prescribed and that observed by each of the three selection methods.

Again, the effect of frequency was expected, given the sloping nature of the average hearing loss for the group, and was not examined further. Post hoc Newman-Keuls testing of the main effect of method revealed that the absolute real-ear gain obtained for aids prescribed by the POGO-II method was significantly greater (p < 0.05) than that obtained for aids prescribed by the NAL-R method. When averaged across frequency, hearing aids fit with the POGO-II procedure yielded 3.6 dB more obtained gain

than instruments fit using the NAL-R method. For the obtained insertion gain, once slight vertical adjustments of the real-ear responses for the POGO-II method were eliminated, there were no differences among the responses obtained with the three selection methods. When the 3.6 dB difference in average absolute gain among methods is considered in light of the typical 30 to 35 dB range of adjustment of the volume control, it is apparent that the actual differences among methods in obtained gain are quite small.

The signed differences between the prescribed and obtained gain values were also determined for each subject, frequency, and selection method. The means and standard deviations of these differences are shown in the bottom panel of Figure 2 for each of the methods and frequencies. There is excellent agreement between prescribed and observed gain for frequencies below 3000 Hz. At higher frequencies (3000 and 4000 Hz), the average observed gain was 5 to 10 dB less than that prescribed by all methods. The signed differences between the obtained and prescribed gain were again subjected to a repeated-measures analysis of variance. Because the high frequencies included in each method are not the same, the data at 3000 Hz were not considered and the data at 2500 Hz for the COX method were included with the data at 2000 Hz available for the other two methods. This analysis revealed that: (a) overall, the signed differences were not significantly different from zero ($F (1,11) = 1.41$, $p > 0.05$); (b) selection method did not have a significant effect on the signed differences ($F (2, 22) = 0.08$, $p > 0.05$); (c) there was a significant effect of frequency on the differences ($F (6, 66) = 5.31$, $p < 0.01$); and (d) the interaction between selection method and frequency was not significant ($F (12, 132) = 1.08$, $p > 0.05$). Newman-Keuls post hoc testing indicated that the effect of frequency on the signed differences between observed and prescribed gain was due to the difference at 4000 Hz being significantly ($p < 0.05$) greater than that at all other frequencies.

Overall, this analysis of the results in Figure 2 suggests that a good match was obtained between prescribed and obtained gain at all but the highest frequency and for all three methods. From a practical standpoint, the prescription made by one method does not appear any easier to realize on the subject than that made by another method. This analysis also indicates

that, although there may be statistically significant differences in the prescribed frequency responses, no significant differences are observed in the obtained responses. The reader should bear in mind, moreover, that electroacoustically flexible BTE instruments with readily modifiable earmolds were used in this study. Yet, despite expending considerable effort adjusting the hearing aid and modifying the earmold to match the prescribed frequency response, significant departures from the prescribed gain were observed in the high frequencies.
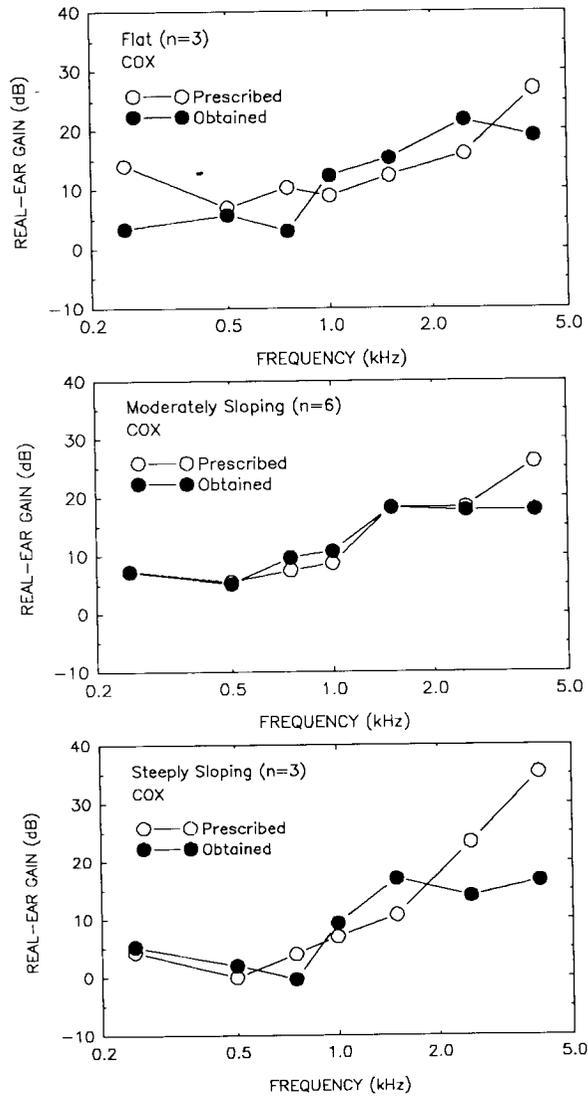
Although the group data suggest equivalence of prescription methods in terms of the ability to match the obtained gain to that prescribed, it is possible that this is not true for all the audiometric configurations included in this study (flat, moderately sloping, and steeply sloping). To examine this possibility, the mean prescribed and obtained insertion gain for each subgroup and each selection method were calculated. These data have been plotted in Figures 3, 4, and 5. Each figure contains the results for one of the prescription methods. For all three methods, the subjects with steeply sloping audiometric configurations (bottom panel of Figures 3, 4 and 5) were the most difficult to fit. For these subjects, the obtained gain at 4000 Hz was 13 to 19 dB less than that prescribed, regardless of selection method. Close matches between obtained and prescribed gain are apparent, however, at the other frequencies for this configuration and for the other two audiometric configurations, regardless of selection method.

The mean speech-recognition scores for the CUNY Nonsense Syllable Test are shown in Figure 6 for aided and unaided listening conditions. For the quiet conditions, all of the aided listening conditions produced scores that were approximately 15 percent better than the unaided condition. There appears to be little difference in speech-recognition performance, however, among the three aided conditions themselves. In noise background, there also appears to be little difference among the aided conditions and also little difference between aided and unaided listening. To analyze these data further, the proportion correct for each condition and subject was transformed using an arcsine transformation (Kirk, 1968). The difference between transformed scores for aided and unaided listening was then computed for each condition and
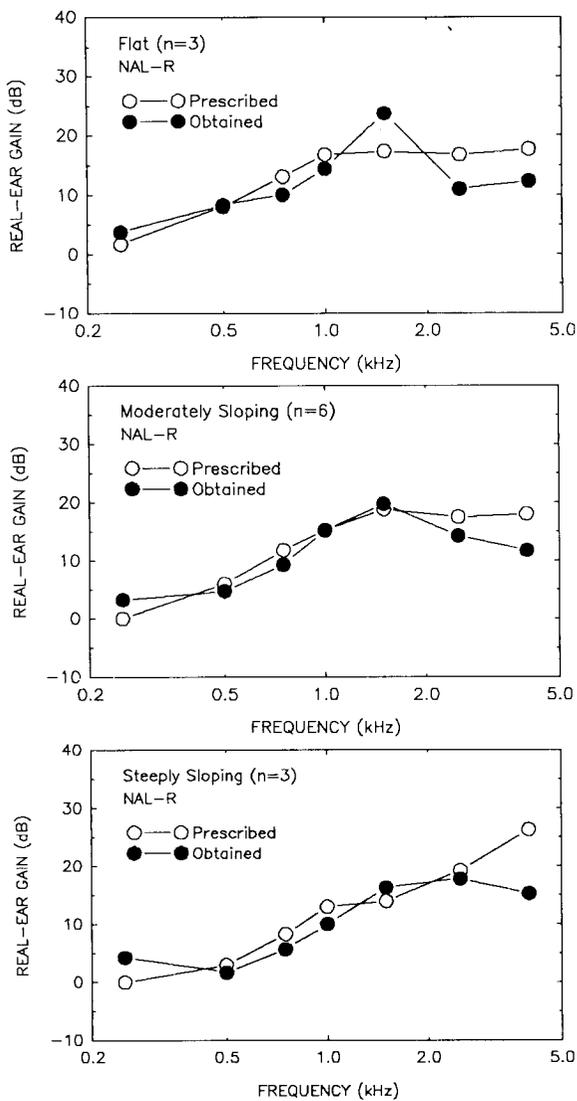
Figure 3 Mean prescribed (*unfilled circles*) and obtained (*filled circles*) insertion gain for three subgroups for the NAL-R selection method. *Top*: subgroup with flat audiometric configuration. *Middle*: subgroup with moderately sloping audiometric configuration. *Bottom*: subgroup with steeply sloping audiometric configuration.
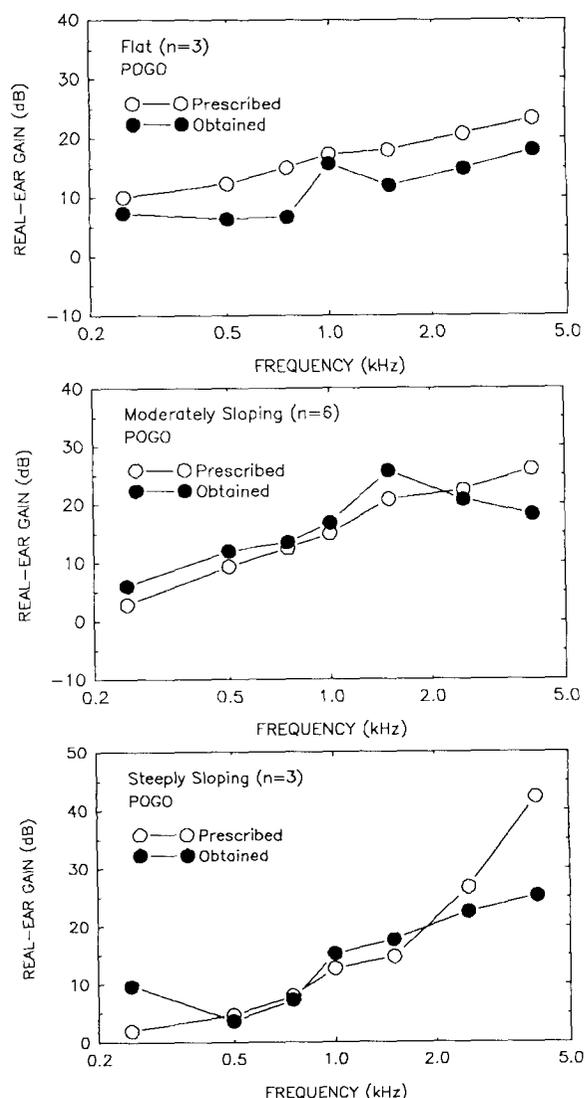
Figure 4 Mean prescribed (*unfilled circles*) and obtained (*filled circles*) insertion gain for three subgroups of subjects for the COX selection method. *Top*: subgroup with flat audiometric configuration. *Middle*: subgroup with moderately sloping audiometric configuration. *Bottom*: subgroup with steeply sloping audiometric configuration.

subject. The differences were then subjected to a repeated-measures analysis of variance which revealed: (a) a significant effect of background $(F (1, 11) = 17.92, p < 0.01)$ with poorer scores observed in noise than in quiet; (b) no effect of selection method $(F (2, 22) = 1.63, p > 0.05)$; and (c) no interaction between background and method $(F (2, 22) = 0.71, p > 0.05)$. Thus, the improvement in speech-recognition performance for aided listening did not vary with selection method; all three methods were equally effective.
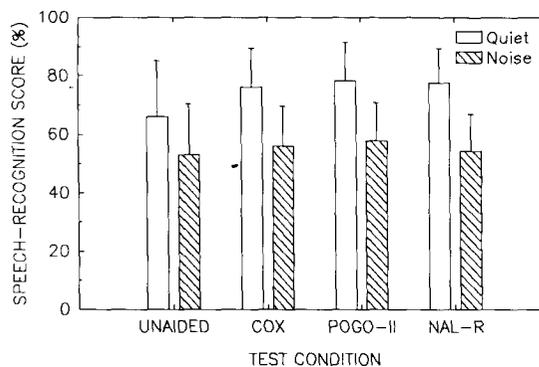
Moreover, selection method did not interact with background, ruling out the possibility that one of the three responses might be better in quiet and one of the other responses better in noise.

Although the group data support the equivalence of the speech-recognition performance obtained with the three selection methods, this may not be the case for the data from individual listeners. Some subjects may perform best with the hearing aid fit with the

Figure 5 Mean prescribed (*unfilled circles*) and obtained (*filled circles*) insertion gain for three subgroups of subjects for the POGO-II selection method. *Top*: subgroup with flat audiometric configuration. *Middle*: subgroup with moderately sloping audiometric configuration. *Bottom*: subgroup with steeply sloping audiometric configuration.

Figure 6 Mean scores on the CUNY NST for unaided and aided listening conditions. Unfilled bars represent quiet listening conditions whereas striped bars represent conditions with the cafeteria-noise background.

POGO-II selection method whereas others may perform best with a hearing aid matching the response prescribed by the NAL-R method, and so on. To evaluate this possibility, all possible paired comparisons of the percent-correct scores were made for the three aided scores in each condition (quiet and noise) and for each subject. Using 95 percent critical differences for a test comprised of 62 items (Thornton and Raffin, 1978), the number of significant differences in

speech-recognition score was determined for each subject. Of the 72 possible paired comparisons (three score pairs in quiet and three in noise for each of the 12 subjects), only one suggested a significant difference in performance. Subject S9 performed significantly poorer in quiet with the hearing aid selected by the COX method than with the instrument selected by the NAL-R procedure. Thus, on both a group *and* an individual basis, the three selection methods yielded essentially identical speech-recognition performance.

The present findings obtained under representative clinical conditions with actual hearing aids fit with clinical procedures confirm earlier results obtained in the laboratory (Sullivan et al, 1988) and previous theoretical calculations using the Articulation Index (Humes, 1986). At least for the three prescriptive selection methods evaluated in this study (NAL-R, COX, and POGO-II), there is little difference among methods in obtained gain characteristics and no difference among methods in aided speech-recognition performance. Although the obtained frequency responses did not differ among methods, it was not because the same hearing aid and earmold were selected for each method. In fact, this was rarely the case. Typically, each prescriptive method resulted in a different instrument and earmold being fit to the patient. These instruments, although not differing significantly in obtained real-ear frequency response, undoubtedly differed in other electroacoustic measures, such as harmonic distortion and internal noise. Despite these likely electroacoustic differences among the aids

selected with each method, the' speech-recognition performance did *not* vary across selection method.

Finally, the emergence of programmable hearing aids will likely reduce the difficulty associated with matching the obtained and prescribed gain. Theoretical analyses (Humes, 1986) and laboratory research with nonwearable instruments (Sullivan et al, 1988), however, both suggest that little difference is expected in aided speech recognition for devices having frequency responses like those prescribed by the three methods evaluated in this study.

REFERENCES

Burney PA. (1972). A survey of hearing aid evaluation procedures. *ASHA*, 14:439-444.

Byrne D. (1988). Recent hearing instrument selection research suggests what? *Hear Instr* 39(12):22-23.

Byrne D, Dillon H. (1986). The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid. *Ear Hear* 7:257-265.

Carhart R. (1946). Selection of hearing aids. *Arch Otolaryngol* 44:1-18.

Cox RM. (1983). Using ULCL measures to find frequency/gain and SSPL90. *Hear Instr* 34(7):17-21,39.

Cox RM. (1988). The MSU hearing instrument prescription procedure. *Hear Instr* 39(1):6-10.

Dubno JR, Dirks DD. (1982). Evaluation of hearing-impaired listeners using a nonsense-syllable test. I. Test reliability. *J Speech Hear Res* 25:135-141.

Humes LE. (1986). An evaluation of several rationales for selecting hearing aid gain. *J Speech Hear Disord* 51:272-281.

Humes LE. (1988). Selecting hearing aids for patients effectively (SHAPE). *Hear J* 14(1):15-18.

Humes LE, Dirks DD, Bell TS, Kincaid GE. (1987). Recognition of nonsense syllables by hearing-impaired listeners and by noise-masked normal hearers. *J Acoust Soc Am* 81:765-773.

Humes LE, Hipskind NM, Block MG. (1988). Insertion gain measured with three probe-tube systems. *Ear Hear* 9:108-112.

Kirk RE. (1968). *Experimental design: procedures for the behavioral sciences.* Belmont, CA: Wadsworth.

Resnick DM, Becker M. (1963). Hearing aid evaluation: a new approach. *ASHA* 5:695-699.

Resnick S, Dubno JR, Hoffnung S, Levitt H. (1975). Phoneme errors on a nonsense syllable test. *J Acoust Soc Am* 58(Suppl. 1):S114.

Schwartz DM, Lyregaard PE, Lundh P. (1988). Hearing aid selection for severe to profound hearing loss. *Hear J* 14(2):13-17.

Shore I, Bilger RC, Hirsh I. (1960). Hearing aid evaluation: reliability of repeated measurements. *J Speech Hear Disord* 25:152-167.

Smaldino J, Hoene J. (1981a). A view of the state of hearing aid fitting practices. Part I. *Hear Instr* 32(1):14-15, 58.

Smaldino J, Hoene J. (1981b). The nature of common hearing aid fitting practices. Part II. *Hear Instr* 32(2):8-11, 43.

Sullivan JA, Levitt H, Hwang J-Y, Hennessey A-M. (1988). An experimental comparison of four hearing aid prescription methods. *Ear Hear* 9:22-32.

Thornton AR, Raffin MJ. (1978). Speech discrimination scores modeled as a binomial variable. *J Speech Hear Res* 21:507-518.

Walden BE, Schwartz DM, Williams DL, Holum-Hardegen LL, Crowley JM. (1983). Test of the assumptions underlying comparative hearing aid evaluations. *J Speech Hear Disord* 48:264-273.