Journal of Speech and Hearing Research, Volume 29, 155–162, June 1986

EFFECTS OF AGING ON RESPONSE CRITERIA IN SPEECH-RECOGNITION TASKS

SANDRA GORDON-SALANT
University of Maryland, College Park

This study examined whether young and elderly listeners exhibit different response criteria on speech-recognition tasks. Young and elderly listeners with normal hearing and with matched mild hearing losses were evaluated on Northwestern University Test No. 6 (Tillman & Carhart, 1966) and the California Consonant Test (Owens & Schubert, 1977) presented at 80 dB SPL and 95 dB SPL. The level of the multitalker babble background was adjusted individually to the signal-to-babble ratio at which the listener achieved 50% criterion performance. Significant differences between the performances of young and elderly listeners were observed on the response bias measure *(B)* but not on the percent-correct or sensitivity [*P(A)*] measures. Elderly listeners exhibited less cautious response criteria than did younger listeners. The implications of these results to communication strategies of elderly listeners are discussed.

Elderly listeners frequently perform more poorly than young listeners on speech recognition tasks. One potential cause of the relative difficulty of elderly listeners is their use of a cautious response criterion during listening tasks. Marshall (1981) hypothesized that an elderly listener with a conservative criterion might refuse to respond to speech stimuli that were unclear. Consequently, an unwillingness to take risks might result in lowered percent-correct recognition scores for elderly listeners. Response criteria (risky vs. conservative) employed by listeners can be derived separately from their sensitivity to a signal using signal detection analysis. Previous efforts (Yanz & Anderson, 1984) to compare young and elderly listeners' performances on speech-recognition tasks were unable to demonstrate criterion differences between groups. The present study attempted to re-examine whether young and elderly listeners exhibit different response criteria on speech-recognition tasks, using a number of listening conditions designed to maximize these differences.

The reported deterioration in speech–recognition performance of older individuals in noise seems to be robust for a variety of different speech materials, response tasks, and types of noise backgrounds. However, age effects on these measures appear to be influenced by several interrelated factors. These include presentation level of the signal (Beattie & Warren, 1983), pure–tone threshold sensitivity of elderly and young listeners (Chung & Mack, 1979; Garstecki & Mulac, 1974; Humes, Schwartz, & Bess, 1979), and perceptual difficulty of the task as influenced by the type of noise background (Garstecki & Mulac, 1974) and signal-to-noise ratio (S/N) (Jokinen, 1973). Studies in which young and elderly subjects were matched for threshold sensitivity have generally confirmed that elderly subjects recognize speech more poorly than do young adults when comparable signal and noise conditions are employed (Dubno, Dirks, & Morgan, 1984; Findlay & Denenberg, 1977; Orchik & Burgess, 1977; Smith & Prather, 1971).

The cause of this age effect is unclear. One possibility is that young and elderly listeners employ different listening strategies. Such listening strategies could influence a listener's response behavior on a perceptual task. Specifically, elderly and young listeners may use a different criterion in judging and responding to a stimulus. In situations with stimulus uncertainty, the elderly listener may be biased toward not responding at all, rather than guessing that a stimulus was present. This type of response behavior would effectively reduce recognition scores.

Differences in response criteria used by young and elderly listeners have been observed on numerous nonauditory and auditory tasks. On nonauditory tasks, elderly subjects tend to behave more cautiously than do younger subjects when given a choice between risky and safe situational alternatives (Wallach & Kogan, 1961). Botwinick (1969) demonstrated that elderly subjects are inclined to avoid risky alternatives when given the avoidance option. This avoidance behavior of elderly subjects contributes to age differences on response-bias measures. When subjects are forced to select risky alternatives, age differences on criterion measures are not observed (Botwinick, 1969).

On pure-tone detection tasks, elderly subjects adopt response criteria that differ from those used by younger listeners. Several studies (Craik, 1966; Potash & Jones, 1977; Rees & Botwinick, 1971) applied signal-detection theory to determine the listener's sensitivity to the signal (i.e., ability to make correct judgments and avoid incorrect ones) independently of response bias (extent to which the subject favors one response over another independent of stimulus events). In all three studies, each listener's sensitivity to the stimuli was equated by adjusting the intensity of the stimuli. The observation intervals consisted of a pure-tone stimulus and white noise ("signal" trials) or white noise alone ("noise" trials). Each type of trial was presented in 50% of the intervals. The subject's response was to indicate "yes" or "no" regarding the presence of the signal (Craik, 1966; Rees & Botwinick, 1971), or to rate the presence of the stimulus during each interval on a 5- or 6-point confidence scale (Craik, 1966; Potash & Jones, 1977). Significant differences in the response criterion measure were reported to

0022-4685/86/2902-0155$01.00/0

occur between young and elderly subjects in all three studies. Specifically, elderly subjects were biased toward responding "noise" (i.e., unwilling to report the presence of the signal), and young subjects were comparatively unbiased. It should be noted that Craik observed similar criteria differences between subject groups in the yes-no task but not in the rating-scale task.

Recently, Yanz and Anderson (1984) investigated response criteria used by young and elderly adults on speech-recognition tasks. Monosyllabic words were presented at 40 dB above spondee threshold (ST) in the presence of broadband noise. The elderly adults had either normal hearing or mild hearing losses. The subjects' tasks included word identification and rating of their identification accuracy on a 6-point confidence scale. For purposes of signal-detection analysis, correct identification responses represented "signal" trials and incorrect identification responses represented "noise" trials. Data analysis revealed that young and elderly listeners were not different in measures of percent correct or decision criterion. However, young subjects were more sensitive than were the older subjects in judging the accuracy of their responses in the more favorable listening condition (+5 dB S/N) but not in the less favorable listening condition (0 dB S/N). Yanz and Anderson concluded that the communication styles of young and elderly listeners cannot yet be distinguished on the basis of decision criterion.

The purpose of the present investigation was to extend and clarify the initial work of Yanz and Anderson (1984). Efforts were made to control for possible interactions between age and hearing loss by employing both normal-hearing listeners and hearing-impaired listeners in two age groups. This permitted evaluation of the contributions of age and hearing loss to speech-recognition measures as well as interactions between these factors. Decision criteria of young and elderly listeners were investigated in an extensive set of listening conditions to examine the generality of such effects. Stimulus materials incorporating both open-set tasks and closed-set tasks were used to evaluate whether the opportunity to avoid a response exaggerates age effects on speech-recognition measures. Open-set tasks provide more opportunity than closed-set tasks for the listener to avoid making a risky decision by not responding. Finally, signal level was varied to determine whether age effects are robust across different listening conditions.

METHOD

Subjects

Subjects were selected from the University of Maryland's Hearing Clinic files on the basis of age and hearing status. Four groups of 10 subjects each participated in this experiment, for a total of 40 subjects. Subjects in Groups 1 and 2 were young adults, ranging in age from 18 to 40 years (M = 26.37). The subjects in Groups 3 and 4 were older adults, ranging in age from 65 to 75 years (M = 67.47). Subjects assigned to Groups 1 and 3 had normal pure-tone thresholds of < 15 dB HL (ANSI-1969) from 250 through 4000 Hz. Subjects assigned to Groups 2 and 4 had sensorineural hearing losses of mild degree, as evidenced by STs ≤ 40 dB HL. The audiometric configurations of the subjects in Groups 2 and 4 were either gradually sloping or sharply sloping. However, every effort was made to match the audiogram of each subject in Group 2 to that of a subject in Group 4. The mean audiograms of the subjects in Groups 2 and 4 are presented in Figure 1. All subjects had word-recognition scores on Northwestern University Test No. 6 (NU6, Tillman & Carhart, 1966) exceeding 85% when the test was presented in quiet. This criterion was used to minimize differences between subject groups on speech-recognition abilities in idealistic conditions. All subjects were free of middle ear disorder, as evidenced by normal tympanograms.

Stimuli

The stimuli included two different standardized tests: one open set and one closed set. The first set of taped

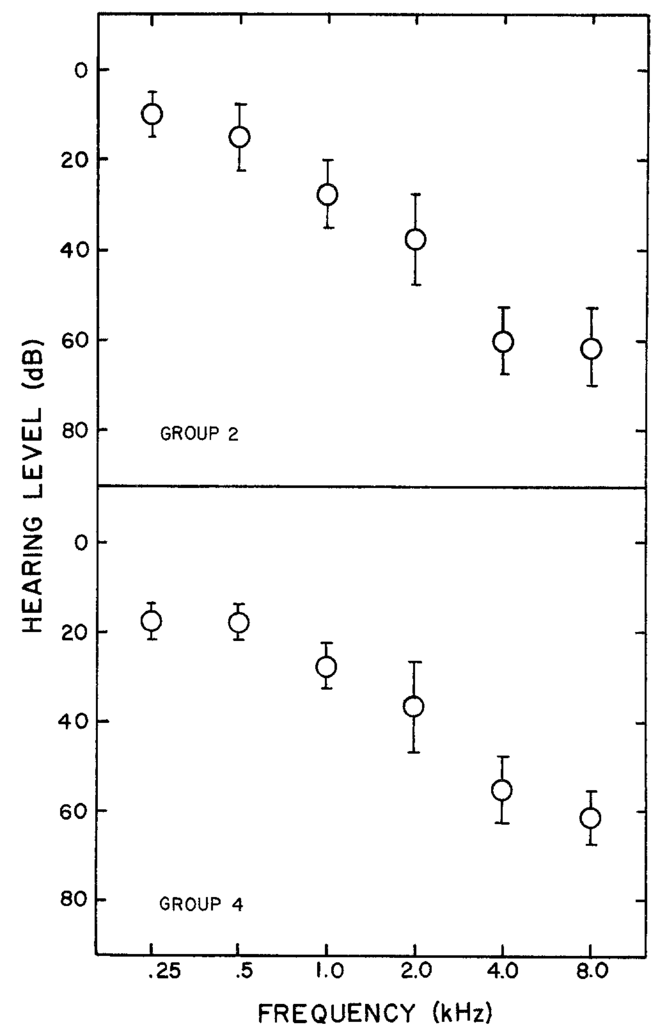

FIGURE 1. Mean pure-tone thresholds of 10 young hearing-impaired subjects (Group 2) and 10 elderly hearing-impaired subjects (Group 4). Vertical bars represent 1 standard deviation.

stimuli consisted of Auditec of St. Louis recordings of NU6 (Tillman & Carhart, 1966), Lists 1A through 4A (open set). The second set of stimuli consisted of Auditec of St. Louis recordings of Lists 1 and 2 of the California Consonant Test (CCT, Owens & Schubert, 1977). The CCT's response format is closed set. In addition, a scrambled randomization of each set of stimuli was prepared for a preliminary test phase.

Noise

The background noise was the 12-talker babble of the Speech Perception in Noise test (SPIN) (Kalikow, Stevens, & Elliott, 1977). This babble was recorded so that fluctuations in level do not exceed ±4 dB of the baseline. The spectrum of the babble is comparable to the long-term average spectrum of speech (French & Steinberg, 1947).

Equipment

The speech stimuli and noise were recorded on separate channels of magnetic tape. During the experiment, the stimuli and noise were played back on a Sony TC-399 tape recorder. The stimuli and noise were separately attenuated (Hewlett-Packard 350D attenuators), mixed (Colbourn audio-mixer amplifier #S82-24), amplified (Crown D75 amplifier), and presented monaurally to the subject via a TDH-49 earphone mounted in an MX41/AR circumaural cushion. The stimuli and noise were presented to the right ear of normal-hearing subjects and to the ear with better threshold sensitivity of hearing-impaired subjects. Speech-spectrum noise was presented to the nontest ear when threshold differences between ears permitted the possibility of crossover. The subject was seated in a double-walled sound-isolated chamber suitable for threshold testing.

Stimuli were calibrated to the level of a peak-equivalent 1000 Hz tone that produced either 80 or 95 dB SPL at the output of the earphones as measured in a 6-cm^3 coupler. The calibration tone was adjusted to equal the peak VU-meter deflection produced by a target word in the carrier phrase. The overall level of the babble was adjusted to produce 50 or 65 dB SPL, as measured in the 6-cm^3 coupler. This created an initial signal-to-babble ratio (S/B) of +30 dB for each signal level condition. Subsequent adjustments in the babble level were made for each subject during testing as described below.

Procedures

Initially, each subject's pure-tone thresholds and tympanogram were obtained. This was followed by a preliminary test phase and an experimental test phase.

The purpose of preliminary testing was to establish the S/B at which the subject attained approximately 50% criterion performance for each of the two stimulus materials presented at each of two test levels. Randomized items from NU6 and CCT were presented on separate runs to the subject at fixed levels of 80 dB SPL and 95 dB SPL. The subject repeated the item perceived for NU6 stimuli and selected the item perceived from a closed set of four choices for CCT stimuli. An adaptive procedure was used to adjust the level of the babble, following procedures described by Dirks, Morgan, and Dubno (1982). The starting level of the babble was 30 dB below the level of the signal (i.e., +30 dB S/B). Babble level was increased in 2-dB steps if the listener correctly identified two out of three stimuli presented at one level. Babble level was decreased in 2-dB steps if the listener exhibited two out of three incorrect responses at one level. The 2-dB adjustment procedure was preceded by an initial search, in which babble was adjusted in 6-dB steps until the first reversal occurred. The 2-dB adjustment procedure continued until eight excursions (changes in direction of babble adjustment) were completed. The S/B for 50% performance was calculated by averaging the median S/Bs of the final four excursions of each run. Thus, measurements of S/Bs were obtained for four conditions, at which two types of stimuli (NU6 and CCT) each were presented at two signal levels (80 and 95 dB SPL). The order of presentation of these four conditions was randomized across subjects.

The experimental test phase consisted of four listening conditions in which two tests (NU6 and the CCT) were each presented at two intensity levels (80 and 95 dB SPL). For each of the two open-set conditions, two NU6 word lists were presented; for each of the two closed-set conditions, one complete CCT word list was presented. Thus, 100 stimuli were presented during each of the four listening conditions, for a total of 400 stimulus presentations. Each list was presented once to ensure that learning of stimulus items did not occur. The babble was presented simultaneously with the stimuli at the level established during the preliminary phase at which criterion performance was achieved. The derived S/B that corresponded to the specific test and stimulus level was used for each listening condition. This procedure of employing the S/B that roughly approximated 50% correct recognition performance was used in accordance with the usual assumption of signal-detection theory that signal and noise events are presented with equal probability. In this experimental paradigm, signal events corresponded to correct word identifications and noise events corresponded to incorrect identification responses.

The subject's task was to write the word perceived for NU6 items and circle one of four choices for CCT items. In addition, subjects were asked to circle one of five categories indicating how sure they were that their responses were correct. The first rating-scale category (++) signified that the subject was absolutely certain that the response was correct; the fifth rating-scale category (−−) signified that the subject was absolutely certain that the response was incorrect. The three rating-scale categories between these two extremes (+, +−, −) represented less degrees of certainty regarding the correctness or incor-

TABLE 1. Mean number of correct words from four subject groups in four listening conditions.

	Listening condition			
	NU6		CCT	
Subject groups	80 dB SPL	95 dB SPL	80 dB SPL	95 dB SPL
Young normal-hearing				
M	55.6	50.9	44.3	54.1
SD	14.0	13.0	10.3	11.9
Young hearing-impaired				
M	56.4	56.9	45.9	55.3
SD	12.0	11.0	10.0	11.4
Elderly normal-hearing				
M	57.0	54.8	50.8	55.1
SD	13.8	13.9	9.8	10.8
Elderly hearing-impaired				
M	55.6	57.0	45.7	54.2
SD	10.0	10.0	12.0	11.0

rectness of the subject's identification response. Examples of the use of each rating-scale category were provided. Subjects were told that approximately 50% of their responses would be correct and 50% of their responses would be incorrect. Consequently, they were informed that they should be using more than one rating-scale category in judging their responses. Indeed, they were encouraged to use all five categories in the rating scale. The procedures employed for deriving hit and false-alarm probabilities from the speech-recognition and rating-scale data are described in the Results section.

The order of listening conditions in the test phase was randomized over subjects. The entire procedure was completed in two sessions of 1.5 hours each, scheduled over a 2-day period. All subjects were paid for their participation in the experiment.

RESULTS

Subjects' responses from each condition were separated into three dependent variables: the total number of correct identification responses *(C)*, the sensitivity to judging response accuracy [*P(A)*], and the response criterion *(B)*. The latter two measures were derived using principles of signal-detection theory, as described below. Each of these three measures was submitted separately for a four-way analysis of variance (ANOVA) in a split-plot factorial design (Kirk, 1968). The SPSS MANOVA procedure (Hull & Nie, 1981) was used to conduct these analyses. In each analysis, the two between-subjects factors were age and hearing status; the two within-subjects factors were test stimulus and signal level. There were two levels of each of these four factors.

The mean number of correct-identification responses for the subjects in each group, for each of the four listening conditions, are presented in Table 1. The raw identification scores of all subjects for all conditions were submitted for ANOVA. The results revealed no significant main effects of age, hearing status, or signal level. A significant main effect of test stimulus was observed ($F(1,36) = 12.28$, $p < .01$). However, there was also a significant interaction between test stimulus and level ($F(1,36) = 14.54$, $p < .01$). Additional interactions were not observed. A simple main-effects analysis of the Test × Level interaction revealed that recognition scores were significantly higher on the NU6 test than on the CCT at 80 dB SPL but not at 95 dB SPL, and that recognition scores were significantly higher at 95 dB SPL than at 80 dB SPL on the CCT.

A number of different sensitivity and response bias measures can be derived from response data. In the present analysis, the nonparametric measures of sensitivity and response bias were derived to avoid assumptions regarding the shape and variance of the underlying distributions of signal and noise. Further, nonparametric measures are preferred when the total number of trials per condition approximates 100 (McNicol, 1972).

The first step in calculating sensitivity and response criterion measures is to derive a set of hit [*P(S/s)*] and false alarm [*P(S/n)*] rates. For the first rating-scale category (high certainty that the response was correct), the hit rate was the total number of correct words when this category was circled (signal responses) divided by the total number of correct words for that condition (signal events). The false-alarm rate was the total number of incorrect words when this category was selected (signal responses) divided by the total number of incorrect words for that condition (noise events). Hit and false alarm probabilities for subsequent categories (less certainty) were obtained by summing the hit rate for the last category (or categories) with the current hit rate, and the false alarm rate for the last category (or categories) with the current false alarm rate. These hit and false alarm rates for each rating scale category can be plotted on a curve with scales of *P(S/s)* and *P(S/n)*. This curve, known as the receiver-operating characteristic (ROC) curve, represents the various degrees of response bias adopted when the stimulus conditions are held constant.[1]

[1]For illustrative examples of ROC curves and discussion of their derivation, the reader is referred to McNicol (1972) and Yanz (1984).

TABLE 2. Mean sensitivity [*P(A)*] values from four subject groups in four listening conditions.

	Listening condition			
	NU6		*CCT*	
Subject groups	*80 dB SPL*	*95 dB SPL*	*80 dB SPL*	*95 dB SPL*
Young normal-hearing				
M	.70	.70	.61	.67
SD	.15	.08	.07	.09
Young hearing-impaired				
M	.75	.72	.65	.66
SD	.05	.03	.05	.05
Elderly normal-hearing				
M	.77	.77	.64	.68
SD	.03	.05	.08	.09
Elderly hearing-impaired				
M	.74	.72	.62	.63
SD	.07	.03	.07	.09

TABLE 3. Mean criterion *(B)* values from four subject groups in four listening conditions.

	Listening condition			
	NU6		*CCT*	
Subject groups	*80 dB SPL*	*95 dB SPL*	*80 dB SPL*	*95 dB SPL*
Young normal-hearing				
M	1.99	2.13	1.57	1.47
SD	.43	.58	.39	.45
Young hearing-impaired				
M	2.00	1.93	1.49	1.49
SD	.61	.40	.46	.40
Elderly normal-hearing				
M	1.68	1.79	1.10	1.11
SD	.75	.71	.46	.44
Elderly hearing-impaired				
M	1.64	1.59	1.26	1.22
SD	.60	.56	.54	.66

The proportion of the area under the ROC curve corresponds to a listener's sensitivity or ability to distinguish between signal and noise events. The sensitivity measure, *P(A)*, can be calculated using McNicol's formula (1972):

$$P(A) = \frac{1}{2} \sum_{i=1}^{N+1} [P_i(S/n) - P_{i-1}(S/n)] \quad [P_i(S/s) + P_{i-1}(S/s)] \qquad (1)$$

where N represents the total number of points, and $P_i(S/n)$ and $P_i(S,s)$ are the coordinates for the ith point on the ROC curve.

The response criterion, B, corresponds to the rating scale category where the sum of the hit rate and false-alarm rate equals 1.0. When B falls between two rating scale categories, it can be interpolated from McNicol's formula (1972):

$$B = \frac{1 - P_l(S/s) - P_l(S/n)}{P_u(S/s) + P_u(S/n) - P_l(S/s) - P_l(S/n)} + C_l, \qquad (2)$$

where $P_l(S/s)$ and $P_l(S/n)$ are the hit and false alarm probabilities for the lower category, $P_u(S/s)$ and $P_u(S/n)$ are the hit and false alarm probabilities for the upper category, and C_l is the size of the lower category in category units.

Sensitivity [*P(A)*] and criterion *(B)* measures were calculated from each subject's identification and rating-scale responses for each listening condition. The average *P(A)* and *B* values for each subject group in each listening condition are shown in Tables 2 and 3, respectively.

Each subject's sensitivity [*P(A)*] values from each condition were arc-sine transformed prior to ANOVA, using the formula $x' = 2\arcsin\sqrt{x}$ (Kirk, 1968). The arc-sine transform is useful to achieve homogeneity of error variance for proportional data (Kirk, 1968). The results of the ANOVA revealed a significant main effect of test stimulus ($F(1,36) = 87.96$, $p < .01$). Additional main effects and interactions were not observed. The main effect of test stimulus reflects higher sensitivity scores for NU6 than for CCT.

The ANOVA of the criterion *(B)* values revealed a significant main effect of age ($F(1,36) = 5.83$, $p < .05$) and a significant main effect of test stimulus ($F(1,36) = 35.79$, $p < .01$). There were no significant interactions. The age effect indicates that younger subjects had significantly higher criterion values than did the elderly subjects for both test materials and both signal levels. The main effect of test stimulus reflects higher criterion values for NU6 stimuli than for the CCT stimuli, which is consistent across all subject groups at both signal levels.

DISCUSSION

Recognition Scores

The S/B adjustment procedure approximating a 50% criterion score was conducted to satisfy the usual assumption that, for each listening condition, the probability of signal and noise presentations would be equal (McNicol, 1972). In addition, this procedure was designed to equate the different subject groups on a baseline measure. If effective, subsequent differences on the sensitivity or response criterion measures observed between groups could not be attributed to different levels of recognition performance. The mean-percent-correct scores shown in Table 1 indicate that word-recognition scores were approximately 50% correct (±7%) for all subject groups in all listening conditions. Significant main effects of age and hearing status were not observed, confirming that the S/B adjustment procedure was effective in equating subjects for percent-correct scores.

One unexpected finding was a significant interaction between test stimulus and level. This interaction reflects that recognition scores were lower on the CCT at 80 dB SPL than they were for all other listening conditions. Thus, although average scores in all conditions approximated 50% correct, our efforts to equate percent correct across all listening conditions were imperfect. The underestimation of criterion score on the CCT presented at 80 dB SPL may be attributed to a high guess rate during the preliminary test phase. Articulation functions derived from normal and hearing-impaired listeners' performances suggest that the CCT is more difficult than is the NU6 at levels below a listener's plateau (Schwartz & Surr, 1979). However, listeners are more inclined to guess when response alternatives are available than when they are unavailable. Thus, the closed-set response format of the CCT may have enabled subjects to attain an artificially low S/B in a difficult listening situation. Consequently, presentation of the CCT at 80 dB SPL during the test phase represented the most difficult listening condition.

Sensitivity

The different subject groups did not exhibit significantly different sensitivity [$P(A)$] scores. In other words, young and elderly subjects, regardless of hearing status, were equal in their ability to judge response accuracy. This finding is consistent with the results of Yanz and Anderson (1984). They reported that elderly subjects were less sensitive than were younger subjects in judging response accuracy at +5 dB S/N, but were equivalent in sensitivity to the young subjects at 0 dB S/N. Thus, in more difficult listening conditions, young and elderly subjects performed equally on the sensitivity measure in that study. The tasks employed in the present investigation were more difficult than those employed in the Yanz and Anderson study, as reflected by observed recognition scores. Average percent-correct scores of young and elderly subjects in Yanz and Anderson's study were 82.9% and 76.0% at +5 dB S/N, respectively, and 64.6% and 59.1% at 0 dB S/N, respectively. In the present investigation, recognition scores of all subjects approximated 50% correct in all conditions. Thus, the current findings support the conclusion that young and elderly subjects cannot be differentiated on the sensitivity measure when difficult listening conditions are employed.

All four subject groups demonstrated significantly higher measures of sensitivity [$P(A)$] for the NU6 test than for the CCT. This was consistent at both presentation levels. Thus, listeners were more able to judge the accuracy of their response when they supplied the response than when they selected the response from a set of alternatives. The CCT employs stimulus items and response foils that are easily confused in ideal listening conditions. In the present experiment, where noise was employed, there is considerable opportunity for extensive confusions. It appears that the availability of foils with a high degree of similarity to the stimulus leads the listener to a false impression of response accuracy. The implication of this finding is that listeners may frequently misjudge their response accuracy when closed-set materials are used in diagnostic and rehabilitative applications. However, the findings reported herein may be specific to the CCT. Further study is warranted to determine whether a listener's sensitivity is comparatively low for closed-set response materials other than the CCT.

Criterion

All subjects exhibited higher numerical values on the criterion measure (B) for the NU6 test than for the CCT test. Relatively high values of B reflect a bias toward noise responses (certain incorrect); relatively low B values reflect a bias toward signal responses (certain correct). The current results indicate that subjects used a more cautious criterion (bias toward noise responses) when responding to NU6 judgments than when responding to CCT judgments. In other words, listeners were more confident in their responses to closed-set materials than to open-set materials. Combining this result with the effects of test material on the sensitivity measure [$P(A)$], it appears that subjects are more confident but less sensitive in judging response accuracy when stimulus alternatives are available than they are when alternatives are unavailable.

Elderly subjects and younger subjects exhibited significantly different response criteria (B values). This age effect was consistent for both normal-hearing and hearing-impaired subject groups, and across all listening conditions. Examination of B values indicates that the older adults had lower numerical B values than did the younger adults. The rating scale task consisted of five response categories. Consequently, a *B* value of 3.0 represents no bias, a *B* value of < 3.0 represents a bias toward signal responses (i.e., "certain correct"), and a *B* value of > 3.0 represents a bias toward noise responses (i.e., "certain incorrect"). We note that the mean *B* values of all subject groups in all listening conditions were < 3.0, suggesting that all subjects were biased toward signal (certain correct) responses. However, the significant difference in *B* values between the young and elderly subjects reflects a greater bias of elderly subjects toward signal (certain correct) responses as compared to the young adults.

The observed age effect is different from previous investigations in which speech perception abilities were evaluated. Yanz and Anderson (1984) found that the response criterion *(B)* values of young and elderly subjects were not significantly different. There are many possible sources that could have contributed to these variant results. For example, methodological differences may account for the discrepant findings. First, the educational backgrounds of subjects in the two studies were different. All subjects in Yanz and Anderson's study were either university graduates (elderly subjects) or undergraduates (young subjects). In the present study, each group was composed of subjects with diverse educational backgrounds. It has been suggested that a subject's education affects his/her verbal skills and confidence, which in turn may influence decision criteria (Yanz & Anderson, 1984). The similarity in response criteria used by young and elderly subjects in Yanz and Anderson's study may have been associated with the similarity and level of their educational background. Thus, the present results may be more representative of the general population. Second, the listening tasks in the current study were more difficult than those used in the Yanz and Anderson study. Recall that average percent-correct recognition scores approximated 50% in the present study but varied between 60 and 83% in the Yanz and Anderson study. The current results suggest that in more difficult listening conditions (as in the present study) older subjects are less likely to identify response errors than are young subjects. A related point is that in Yanz and Anderson's study, the assumption of equal probability of signal and noise trials was not met. This may have influenced the response criteria employed by listeners, or may have reduced the reliability of point estimates for the ROC curve (Ogilvie & Creelman, 1968).

The expectation that criterion differences exist between young and elderly subjects was met in the present experiment. However, the direction of observed differences was opposite from that reported in pure-tone detection experiments. Previous studies found that elderly subjects were more conservative than were young adults in response behavior; the present experiment showed that elderly subjects were less conservative than were the younger subjects. Differences in the signal presentation and judgment task between a pure-tone detection experiment and a speech-recognition experiment may have contributed to these inconsistent results. In the pure-tone detection paradigm, an externally generated signal or noise is presented at a level approximating the subject's threshold. The subject's task is to indicate the degree of confidence in the presence of the signal. In the speech-recognition paradigm, a signal event is a correct word-recognition response from the subject, and a noise event is an incorrect word-recognition response from the subject. The subject's task is, therefore, to identify the stimulus item and then to indicate the degree of confidence in response accuracy. Thus, the pure-tone detection task evaluates a listener's confidence in the presence of an externally controlled event, whereas the speech-recognition task evaluates a listener's confidence in his or her own identification response. The findings to date suggest that the response bias of elderly listeners is task dependent: elderly listeners are relatively conservative in committing themselves to a signal response for externally controlled events but are relatively confident in the accuracy of their own responses.

Further support of the notion that the task influences observed response behavior can be derived by examining the overall bias of all subjects on pure-tone versus speech-recognition tasks. Comparisons can be made among studies in which the same statistic for response criterion, *B*, was calculated. Potash and Jones (1977) reported average criterion *(B)* values for young and elderly subjects ranging from 3.18 to 5.01, for various listening conditions. Values > 3.0 indicate that all subjects were biased toward noise responses, although elderly subjects were significantly more biased than were younger subjects. In the present experiment and that of Yanz and Anderson (1984), average *B* values of all subject groups in all listening conditions were < 3.0, representing a bias toward signal responses. These preliminary observations suggest that the direction of the response bias is task-dependent, and that elderly subjects exhibit a more exaggerated bias than do younger subjects in the direction of the overall response bias for a particular paradigm. Further research should be directed toward determining whether the judgment task differentially influences criterion effects on other measures.

The major finding of this study is that elderly subjects employ a less conservative criterion than do younger subjects when judging the accuracy of their responses on word-recognition tasks. This result does not support the hypothesis of Marshall (1981) that poor performance on speech-recognition tasks by elderly subjects may be attributed in part to a cautious response bias. Indeed, observed speech-recognition scores of elderly listeners may be inflated relative to their sensory capabilities because of a liberal criterion. That is, a high degree of confidence in one's own responses may encourage the elderly person to guess extensively during speech-recognition tasks. This guessing behavior could result in chance improvement in percent-correct scores. There are also implications of the elderly listener's risky criterion in daily communication situations. The picture emerging is that elderly listeners may misunderstand the spoken message but respond as if they have understood the message. Consequently, their own response to the message is inappropriate and communication inefficiency is perpetuated. Signal-detection techniques may prove valuable in training the elderly person to be less biased in communication tasks. Rehabilitation-training sessions can employ different rewards and penalties to manipulate the listener's criterion until it approximates unbiased response behavior. Analysis techniques similar to those used in the present experiment can confirm whether or not the reward structure effectively altered the subject's bias. Additional research is needed to clarify whether manipulating the response bias of elderly subjects can be accomplished with these methods; and whether communicative efficiency improves with this type of rehabilitative program.

ACKNOWLEDGMENTS

The author would like to thank Sylvia Campbell for her assistance in the data analysis, and Peter Fitzgibbons for his valuable comments on an earlier draft of this manuscript.

REFERENCES

American National Standards specifications for audiometers. (ANSI S3.6-1969). (1970). New York: ANSI.

BEATTIE, R. C., & WARREN, V. (1983). Slope characteristics of CID W–22 functions in elderly hearing-impaired listeners. *Journal of Speech and Hearing Disorders, 48,* 119–127.

BOTWINICK, J. (1969). Disinclination to venture response versus cautiousness in responding: Age differences. *Journal of Genetic Pyschology, 115,* 55–62.

CHUNG, D. Y., & MACK, B. (1979). The effect of masking by noise on word discrimination scores in listeners with normal hearing and with noise-induced hearing loss. *Scandinavian Audiology, 8,* 139–143.

CRAIK, F. I. M. (1966). The effects of aging on the detection of faint auditory signals. *Proceedings of the Seventh International Congress of Gerontology, 6,* 145–147.

DIRKS, D. D., MORGAN, D. E., & DUBNO, J. R. (1982). A procedure for quantifying the effects of noise on speech recognition. *Journal of Speech and Hearing Disorders, 47,* 114–122.

DUBNO, J. R., DIRKS, D. D., & MORGAN, D. E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *Journal of the Acoustical Society of America, 76,* 87–96.

FINDLAY, R. C., & DENENBERG, L. J. (1977). Effects of subtle mid-frequency auditory dysfunction upon speech discrimination in noise. *Audiology, 16,* 252–259.

FRENCH, N. R., & STEINBERG, J. C. (1947). Factors governing the intelligibility of speech sounds. *Journal of the Acoustical Society of America, 19,* 90–114.

GARSTECKI, D. C., & MULAC, A. (1974). Effects of test material and competing message on speech discrimination. *Journal of Auditory Research, 3,* 171–178.

HULL, C. H., & NIE, N. H. (1981). *SPSS Update 7–9.* New York: McGraw-Hill.

HUMES, L. E., SCHWARTZ, D. M., & BESS, F. H. (1979). Two experiments on subtle mid-frequency hearing loss and its influence on word discrimination in noise-exposed listeners. *Audiology, 18,* 307–319.

JOKINEN, K. (1973). Presbyacusis VI. Masking of speech. *Acta Otolaryngologica, 76,* 426–430.

KALIKOW, D. N., STEVENS, K. N., & ELLIOTT, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61,* 1337–1351.

KIRK, R. E. (1968). *Experimental design: Procedures for the behavioral sciences.* Belmont, CA: Brooks/Cole Publishing.

MARSHALL, L. (1981). Auditory processing in aging listeners. *Journal of Speech and Hearing Disorders, 46,* 226–240.

MCNICOL, D. (1972). *A primer of signal detection theory.* London: George Allen & Unwin Ltd.

OGILVIE, J. C., & CREELMAN, C. D. (1968). Maximum-likelihood estimation of Receiver Operating Characteristic curve parameters. *Journal of Mathematical Psychology, 5,* 377–391.

ORCHIK, D. J., & BURGESS, J. (1977). Synthetic sentences identification as a function of the age of the listener. *Journal of the American Audiology Society, 3,* 42–46.

OWENS, E., & SCHUBERT, E. D. (1977). Development of the California Consonant Test. *Journal of Speech and Hearing Research, 20,* 463–474.

POTASH, M., & JONES, B. (1977). Aging and decision criteria for the detection of tones in noise. *Journal of Gerontology, 32,* 436–440.

REES, J. N., & BOTWINICK, J. (1971). Detection and decision factors in auditory behavior of the elderly. *Journal of Gerontology, 26,* 133–136.

SCHWARTZ, D. M., & SURR, R. K. (1979). Three experiments on the California Consonant Test. *Journal of Speech and Hearing Disorders, 44,* 61–72.

SMITH, R. A., & PRATHER, W. F. (1971). Phoneme discrimination in older persons under varying signal-to-noise conditions. *Journal of Speech and Hearing Research, 14,* 630–638.

TILLMAN, T. W., & CARHART, R. (1966). *An expanded test for speech discrimination utilizing CNC monosyllabic words. Northwestern University Auditory Test No. 6.* Brooks Air Force Base, TX: USAF School of Medicine.

WALLACH, M. A., & KOGAN, N. (1961). Aspects of judgements and decision making: Interrelationships and changes with age. *Behavioral Science, 6,* 23–36.

YANZ, J. L. (1984). The application of the Theory of Signal Detection to the assessment of speech perception. *Ear and Hearing, 4,* 64–71.

YANZ, J. L., & ANDERSON, S. M. (1984). Comparison of speech perception skills in young and old listeners. *Ear and Hearing, 5,* 134–137.

Received August 5, 1985
Accepted October 24, 1985

Requests for reprints should be sent to Sandra Gordon-Salant, Department of Hearing and Speech Sciences, University of Maryland, College Park, MD 20742.