Impacts of age on memory for auditory intensity^{a)}

Frederick J. Gallun,^{b)} Anna C. Diedesch,^{c)} and Robertson Beasley^{d)}

VA RR&D National Center for Rehabilitative Auditory Research (NCRAR), Portland VA Medical Center,
3710 SW US Veterans Hospital Road, Portland, Oregon 97239

(Received 6 October 2011; revised 7 June 2012; accepted 12 June 2012)

It is hypothesized that older listeners are more likely than younger listeners to be impaired when asked to make intensity judgments about target tones embedded in rapidly presented auditory sequences. This study examined this hypothesis by asking listeners ranging in age from 19 to 74 yr to make judgments of intensity based on narrowband noise bursts varying in frequency and intensity. In two experiments, listeners made intensity judgments of target bursts alone or embedded in sequences of bursts. In the first experiment, one of four fixed sequences was presented and had to be identified. In the second experiment, pre- or post-trial bursts acted as cues that identified the frequency of the target burst in the sequence. In both experiments, intensity discrimination thresholds for single bursts were good predictors of performance with sequences and were little affected by age. Significant negative relationships between age and accuracy were observed when single sequences had to be identified or a post-trial cue was used, but no age effects were apparent when a pre-trial cue was used. These data are interpreted as being consistent with previous suggestions that the aging process results in a decline in auditory memory capacity and/or internally generated selective attention. [<http://dx.doi.org/10.1121/1.4731235>]

PACS number(s): 43.66.Ba, 43.66.Fe, 43.66.Mk, 43.66.Sr [LD]

Pages: 944–956

I. INTRODUCTION

Sensitivity to changes in frequency and intensity over time can be described as sensitivity to spectrotemporal patterns (Chi *et al.*, 1999), which in turn can be considered, along with sensitivity to the durations of the pattern elements, as the basic building blocks of most behaviorally relevant auditory stimuli. Many studies that have examined the ability of listeners to process such patterns have determined that the presence of irrelevant frequencies and/or intensities at times adjacent to the time at which the target occurs can interfere with both the detection and discrimination of target stimuli (Watson *et al.*, 1975; Spiegel and Watson, 1981; Leek and Watson, 1984; Espinoza-Varas and Watson, 1986; Tremblay *et al.*, 2005; Vachon and Tremblay, 2005, 2006, 2008; Watson, 2005; Vachon *et al.*, 2007; Cousineau *et al.*, 2009; Gallun, 2010). This is the case even when frequency overlap between the distractors and the target is minimized (Kidd *et al.*, 2008). Auditory processing among older listeners is assumed to be impaired primarily by presbycusis, which has the potential to interfere with basic detection processes and may yield supra-threshold deficits as well. There are also many examples in the literature of the negative influences of aging on temporal processing (Trainor and Trehub, 1989; Alain *et al.*, 1996; Fitzgibbons and Gordon-Salant, 1996, 1998, 2001;

Wingfield *et al.*, 1999; Salthouse, 2000; Gordon-Salant *et al.*, 2006; Fitzgibbons *et al.*, 2007). Trainor and Trehub (1989) and Alain *et al.* (1996) both presented data that were interpreted as evidence that while auditory stream segregation ability is intact in older adults, the ability to make judgments about the elements of the streams is impaired.

The goal of this study was to examine the impacts of aging on the ability to make judgments of intensity in temporal sequences. The design is based on the work described by Gallun (2010), in which three narrowband noise bursts were presented in isolation, simultaneously, or sequentially with a delay of 50 or 200 ms. Sequential conditions were presented such that the center frequencies of the bursts increased over time. In every condition, the task of the listener was to discriminate the intensities of the burst or bursts comprising a baseline stimulus (or stimulus sequence) presented 500 ms before the target stimulus. Target bursts were either the same intensity as the baseline stimulus or were incremented by an adaptively-varied amount. Using this approach, Gallun (2010) found that sequential presentation of the bursts led to significantly higher thresholds than in the single-burst or simultaneous cases. There was no reliable effect of age on intensity discrimination for single bursts. Although the highest increment thresholds in the sequential task were for older listeners, on average the five younger listeners (aged 28–36 yr) did not perform significantly differently from the five older listeners (aged 51–60 yr).

The results of Gallun (2010) are consistent with data collected by Tremblay and colleagues (Tremblay *et al.*, 2005; Vachon and Tremblay, 2006, 2008) suggesting that an important component of the ability to process auditory sequences involves the rapid encoding of sensory information in a form that supports later comparisons among stored representations. According to this explanation, the reduced accuracy of listeners in the sequential conditions of Gallun (2010) may have

^{a)}Portions of this work were presented at the 32nd and 33rd Midwinter Meetings of the Association for Research in Otolaryngology.

^{b)}Author to whom correspondence should be addressed. Also at: Dept. of Otolaryngology/Head and Neck Surgery, Oregon Health and Science University, Portland, OR 97239. Electronic mail: Frederick.Gallun@va.gov

^{c)}Present address: Department of Speech and Hearing Sciences, University of Washington, 1417 N.E. 42nd St., Seattle, WA 98105.

^{d)}Present address: Department of Psychology, University at Buffalo, The State University of New York, Park Hall 337, Buffalo, NY 14260.

represented not an inability to extract the intensity information from each burst but rather a failure to use information about the intensity of each burst in a later analysis. Because the age effects were smaller in this task than in similar frequency-discrimination tasks reported in the literature (e.g., [Alain et al., 1996](#)), it was further hypothesized that the size of the age effect might depend on the degree to which such sensory transfer was necessary for success on the task or on the number of elements that could be stored in such a short-term buffer. Alternatively, it might be the case that age effects observed in temporal and pitch judgments do not occur for intensity judgments. To examine these questions, two further experiments were developed that required the identification of combinations of frequency and intensity for two of the sequence elements.

Based on a wide range of studies on auditory masking (both forward and backward), [Cowan \(1984, 1988, 1995\)](#) has argued there are two distinct methods of storing auditory information for subsequent analysis. The first is an initial brief sensory store that decays (or becomes otherwise inaccessible) after between 50 and 150 ms. The second storage method is a non-sensory representation that persists for 4–10 s and can be used for comparisons occurring at longer time scales. For further discussions of the architecture of auditory memory, see [Cowan \(1984, 1988, 1995\)](#), [Demany and Semal \(2008\)](#), and [McKeown and Wellsted \(2009\)](#). To ensure that the effects being examined in the studies reported in the following text were more likely to be related to memory processes than to masking effects due to overlap of stimuli in the brief sensory store, all stimuli were presented with temporal separations of at least 150 ms. In addition, and as noted in the following text, several design choices were made to establish baseline behavioral data for a concurrent electrophysiological examination, data from which are not included. Most importantly, the stimuli were “frozen” to allow repeated presentations of acoustically identical tokens, and the probability of a standard stimulus was 55% to encourage the elicitation of auditory evoked potentials to deviant stimuli.

II. EXPERIMENT ONE: IDENTIFICATION OF FIXED SEQUENCES

A. Listeners

Forty-two listeners participated, ranging in age from 25 to 74 yr (mean 49). Of these, 23 were younger than 50 yr old (18 males and 5 females) and 19 were 50 yr or older (14 males and 5 females). Hearing sensitivity was normal or only mildly impaired for all participants: thresholds averaged across octave frequencies between 0.25 and 8 kHz ranged from 3.3 to 33.3 dB HL (mean of 14.4 dB HL). Older listeners had consistently higher average threshold values than did the younger listeners ($t_{40} = 2.57$, $P = 0.014$), but the average threshold for the older listeners (17.3 dB HL) was only 5 dB higher than for the younger (12.0 dB HL). The greatest differences across groups were in the higher frequencies (4 and 8 kHz), where the thresholds for the older listeners were 7–15 dB higher on average than were those of the younger listeners. Losses were almost always symmetri-

cal across ears (no differences greater than 20 dB at any frequency), but four of the impaired listeners had relatively normal thresholds for either the right or left ear. All listeners were paid for their participation and all procedures were reviewed and approved by the Institutional Review Boards of both the Portland VA Medical Center and the Oregon Health and Science University.

B. Stimuli

Stimuli were 50-ms noise bursts presented in three frequency ranges: low (0.40–0.56 kHz), middle (1.62–2.40 kHz), or high (4.08–6.10 kHz). One noise token was generated for each of the frequency ranges using MATLAB (Mathworks). For each token, a frequency-domain vector was first created with components spaced 20 Hz apart and stretching between 0.40 and 6.10 kHz. Amplitude for a given token was set to unity amplitude in the appropriate range to generate a “low,” “middle,” or “high” burst, and amplitudes were set to zero outside of that range. A phase vector was then generated with values that were randomized across components, and the amplitude and phase vectors were then combined before performing an inverse FFT to create time-domain signals. Bursts were then multiplied with a 50-ms duration amplitude envelope containing 5 ms $\cos^2$ onset and offset ramps. The same three noise bursts were scaled digitally to produce the stimuli for each subject across the various conditions (described in the following text). An Edirol UA-101 (Roland Corp.) converted the digital signal, which was sampled at 48 kHz with 24-bit resolution, into an analog signal that was amplified using a custom buffering amplifier and transduced with ER-2 (Etymotic Research) insert earphones for diotic presentation.

C. Procedures

The procedures were designed to allow rapid testing of a large number of psychoacoustically inexperienced listeners in a challenging test of intensity-sequence identification. It was important to be able to quickly train subjects and obtain baseline performance to then proceed to an electrophysiological experiment involving the sequence identification task. To achieve these goals, a set of three tasks was developed that provided both psychophysical training and familiarization with the stimuli. The techniques were designed to build from simple detection, to intensity discrimination of single stimuli, and finally to identification of sequences based on intensity differences. In the first two tasks, two intervals were used, and listeners were given trial-by-trial feedback, while in the final task, only a single interval was presented, and listeners were given feedback during a training phase and then tested without being given feedback. The testing without feedback was identical to the testing that would occur during the electrophysiological testing (results of which are not included here).

In the first part of the experiment, detection thresholds (DTs) were obtained. The three frozen noise bursts (low, middle, and high) were presented as single bursts in blocked adaptive tracks. DTs were based on the average of three adaptive tracks for each noise burst. On each trial of the track, two intervals were presented to the listener, one, both,

or neither of which could contain a noise burst. If the noise burst was present on both intervals (25% of the trials) or neither interval (25% of the trials), the listener was correct if they responded by pressing a button marked “same,” while if the noise burst was present on only one interval (50% of the trials), the listener was correct if they pressed the button marked “different.” Each interval was indicated by a 500-ms visual marker, and the intervals were separated by 500 ms of silence. Feedback was given after each trial. RMS level of the noise burst was increased or decreased from an initial level of 30 dB SPL in a three-down/one-up adaptive tracking procedure (Levitt, 1971). The procedure lasted for 10 reversals; the RMS level was raised or lowered in 4-dB steps until the first reversal, 2-dB steps until the second reversal, and 1-dB steps thereafter. The last eight reversals were averaged to give a DT estimate and estimates were collected for each of three tracks.

This method differs from an objective tracking procedure in that there are trials in which only one type of stimulus is presented, potentially allowing shifts in decision criterion to impact thresholds. It was preferred because the sequential task was designed to be an identification task, and it was hoped that this method would familiarize listeners with the need to make decisions based on their memories of the intensity of the signal. Although this method requires adjusting the level of the signal on trials in which silence was presented on both intervals, this issue did not seem to lead to particularly variable tracks, and the thresholds measured were very similar to what would have been predicted by pure tone thresholds measured with traditional audiometric methods.

In the second part of the experiment, increment thresholds (ITs) were obtained. As in the first part of the experiment, the three noise bursts were presented as single bursts in blocked sections, and three tracks were averaged to give the IT for each burst. For the IT task, bursts were presented in each of the two observation intervals, and the task of the listener was to determine whether or not the intensities of the two noise bursts were the same or different by pressing buttons marked “same” and “different.” Each interval was indicated by a 500-ms visual marker, and the intervals were separated by 500 ms of silence. Intensities were the same on 50% of the trials and different on 50% of the trials, and feedback was given after each trial. The baseline intensity of the noise bursts for each interval was set to a pedestal of 30 dB above that subject’s DT for that noise burst, or 30 dB sensation level (SL). Initially, an increment of 10 dB was added to this pedestal on trials in which the intervals differed. Using a three-down/one-up tracking procedure, this increment was raised or lowered in steps of 4 dB until the first reversal, 2 dB until the second reversal, and 1 dB thereafter. The last eight reversals were averaged to determine the threshold for that track, and three estimates were averaged to produce the IT. This procedure is also not an objective tracking method as there are trials in which no increment is presented, and listeners are required to answer “same” to be correct. The low increment thresholds and consistent performance obtained with this method suggests that the listeners were able to respond consistently despite the presence of non-signal trials.

In the third part of the experiment, the ability of listeners to identify the members of a set of four three-burst sequences was measured using a single-interval, four-choice procedure, as indicated by the schematic shown in Fig. 1. In Gallun (2010), it was found that sensitivity to intensity changes in the low-frequency burst was much greater than sensitivity to changes in the middle and high bursts. For this reason, only changes in the middle and high bursts were used in this experiment. In addition, a two-burst change condition was added to examine the possibility that listeners were using an overall level comparison strategy rather than attending to specific frequencies. If that was the case, then identification performance would be best when listeners were presented with a signal with no incremented bursts or with two incremented bursts, and confusions would be most common between stimuli where only a single burst was incremented.

On each trial, one of four possible sequences of the three 50-ms noise bursts was presented in ascending frequency order, each separated by 150 ms of silence. On 55% of the trials, there was *no change* in the intensity of the stimuli (all were presented at 30 dB SL), and subjects were to respond by pressing the button marked “no change.” On 15% of the trials, the *middle noise burst* was incremented by 15 dB and thus presented at 45 dB SL. Low and high were at 30 dB SL, and the correct response was to select the button marked with the word “middle.” Another 15% of the trials contained only the *high burst* incremented by 15 dB (presented at 45 dB SL). Low and middle bursts were presented at 30 dB SL, making the correct response to choose the button marked “high.” Finally, the remaining 15% of trials featured a

FIG. 1. Schematic description of the low, middle, and high frequency stimuli presented in Experiment One. Larger squares represent greater intensity, and frequency range is indicated by the color of the square as indicated by the legend at the top. Listeners were trained to label sequences with the four descriptors shown to the right of the sequences.

FIG. 2. Detection thresholds (a) and increment thresholds (b) for older (gray circles) and younger listeners (black squares) in Experiment One. Error bars indicate two standard errors of the mean (SE).

sequence in which the low noise burst was presented at 30 dB SL, and *both the middle and high bursts* were presented at 45 dB SL for which the correct response was to choose the button marked “both.” After doing two practice blocks of 100 trials with a 20 dB increment in which feedback was presented, subjects completed three blocks of 100 trials with the 15 dB increments and no feedback. The entire experiment took about 90 min.

D. Results

Detection thresholds for the three noise bursts were fairly similar within each age group but differed for the older and younger groups. For the younger group, DTs were 8.1, 7.8, and 8.9 dB SPL for the low, middle, and high bursts, respectively, whereas for the older group, DTs were 12.1, 10.1, and 14.1 dB SPL. Average DTs for the two groups are plotted in Fig. 2(a) with younger participants indicated by the black squares and older participants by the gray circles. A mixed-models ANOVA with age group as a between-subjects factor and frequency as a repeated-measures factor found no significant effect of frequency ($F_{2,80} = 2.78$, $p = 0.07$), a significant effect of age group ($F_{1,40} = 4.31$, $p = 0.04$), and no significant interaction ($F_{2,80} = 0.95$, $p = 0.39$). As age and pure-tone thresholds were significantly correlated, especially for the higher frequencies, stepwise multiple linear regression was used to determine whether the effect of age group reflected differences in pure-tone thresholds or if age of the listener was still associated with performance after accounting for hearing loss. In all cases, pure-tone thresholds were strong predictors, and age explained no significantly greater proportion of the variance. For the low burst, DT was best predicted by pure-tone thresholds at 500 and 1000 Hz ($R^2 = 0.63$, $p < 0.001$), DT for the middle burst was best predicted by pure-tone thresholds at 2000 Hz ($R^2 = 0.73$, $p < 0.001$), and DT for the high burst was best predicted by pure-tone thresholds at 4000 and 8000 Hz ($R^2 = 0.70$, $p < 0.001$). Additional variables did not significantly improve the model predictions. These patterns are consistent with the frequency content of the noise bursts

and support the use of equal SL stimuli to improve the chances of distinguishing the effects of age from the effects of hearing loss.

Increment thresholds for the three bursts differed with frequency but were fairly similar for the older and younger listeners. For the younger listeners, ITs were 3.5, 5.1, and 5.7 dB for the low, middle, and high bursts, respectively, while for the older listeners, the same bursts produced thresholds of 3.7, 4.7, and 5.3 dB. Average ITs for the two groups are plotted in Fig. 2(b). A mixed-models ANOVA was conducted with age group as a between-subjects factor and frequency as a repeated-measures factor. A significant effect of frequency was observed ($F_{2,80} = 33.93$, $p < 0.001$), but the effect of age group was not significant ($F_{1,40} = 0.07$, $p = 0.79$). There was no significant interaction between age group and stimulus frequency ($F_{2,80} = 0.90$, $p = 0.41$).

Table I shows the results of the identification task, displayed as the stimulus-response matrix across stimuli for the younger and older subjects. The value in each cell represents the proportion of times each response was chosen when a given stimulus was presented. The proportion of correct responses (indicated by the cells with darker lines surrounding them) is also shown in Fig. 3(a). Table I also reveals the proportion of trials on which a given stimulus was chosen, but one of the other three stimuli had been presented (false alarm rates). Examination of these patterns reveals an uneven pattern of errors with the incorrect responses for the middle and high stimuli falling primarily in the no change category, and most of the incorrect responses for the both stimulus falling in the middle and high categories. This pattern suggests that simply examining the hit rate (proportion correct or PC) is not revealing the pattern of results and may be overestimating sensitivity, particularly for the no change stimulus.

The usual method for obtaining a bias-free measure of sensitivity would be to convert the hits and false alarms into d' . The problem is that d' is defined as the distance between two distributions not the four distributions associated with the four possible stimuli presented in this experiment. While there are a number of solutions that have been proposed for converting between percent correct and d' for multiple alternatives (Hacker and Ratcliff, 1979; Green and Dai, 1991; Macmillan and Creelman, 2005), the majority of these

TABLE I. Stimulus-response matrix for the four stimuli presented in the identification task in Experiment One, for the younger ($n = 23$) and older ($n = 19$) listeners.

Response	Group	Stimulus			
		Middle	High	Both	No Change
Middle	Younger	0.59	0.03	0.15	0.02
	Older	0.40	0.08	0.20	0.05
High	Younger	0.06	0.72	0.20	0.01
	Older	0.09	0.56	0.32	0.04
Both	Younger	0.12	0.09	0.60	0.01
	Older	0.14	0.17	0.41	0.04
No Change	Younger	0.23	0.17	0.05	0.95
	Older	0.38	0.19	0.07	0.87

FIG. 3. Accuracy in labeling of sequences in Experiment One by younger (black squares) and older listeners (gray circles). Error bars indicate 2 SE.

approaches assume that the observer is unbiased. This is not useful in this case because the question here is precisely whether or not the presumed (and reasonable) bias toward responding “no change” has impacted the results. Accordingly, the data were reanalyzed using a method first suggested by Eguchi and Hirsh (1969) and used more recently by Leek *et al.* (1987). The basic logic of the Eguchi and Hirsch approach is that each judgment can be viewed as a discrimination between the presented stimulus and the other possible stimuli. Thus trials on which a given stimulus is correctly identified are coded as a “hit,” and trials on which the observer reports that stimulus but one of the other stimuli was actually presented are coded as a “false alarm.” Details of the calculation and further analysis of the method are presented in the Appendix.

Average d' values in the identification task were calculated for each listener using the method of Eguchi and Hirsch (1969) and then averaged across the younger and older listener groups. The resulting values are plotted in Fig. 3(b). The reduced performance for the older listeners is apparent in both the PC and d' metrics. Sensitivity to the middle stimulus as estimated with d' [plotted in Fig. 3(b)] revealed values of 2.01 for the younger listeners and 1.12 for the older listeners. Sensitivity to the high stimulus was 2.33 for the younger listeners and 1.60 for the older. Sensitivity to the both stimulus was 2.05 for the younger listeners and 1.23 for the older. The no change stimulus showed greater sensitivity among all listeners, but the same pattern of differences between groups with a sensitivity of 2.96 for the younger listeners and 2.19 for the older. Overall, the d' difference in sensitivity between groups range from 0.74 for the high stimulus to 0.89 for the middle stimulus.

A mixed-measures analysis of variance showed that there were identical patterns for both PC and d' . For the d' values, the effect of the repeated-measures factor of stimulus was significant, even after applying a Greenhouse–Geisser correction to the degrees of freedom to account for a viola-

tion of the assumption of data sphericity ($F_{2.5,120} = 54.68$, $p < 0.001$), a significant effect of age group ($F_{1,40} = 8.29$, $p = 0.006$), and no significant interaction ($F_{2.5,120} = 1.01$, $p = 0.35$). In summary, the d' analysis presents precisely the same picture of the relative performance of the two groups across the four stimuli as does the percentage correct. This is useful as it confirms that the differences between the groups are not due simply to a difference in the biases of the older and younger listeners but truly reflect their ability to discriminate among the four stimuli.

To determine the extent to which age *per se*, rather than detection or increment thresholds, was responsible for the differences between groups, stepwise regression was used to predict the d' values for each of the four stimuli. DTs, ITs, and current age of the listener were all included as potential predictors. For all four d' measures, 30%–40% of the variance could be accounted for by a model including only age and IT for the high frequency burst ($p < 0.001$ in all cases). The IT value was always the best predictor, accounting for 26% of the variance for the “no change” and “mid” conditions, 19% in the “high” condition, and 34% of the variance in the “both” condition. Age accounted for an additional 13% of the variance in the “no change” condition and 11% in the remaining three conditions. DTs did not significantly improve the predictions of the model.

E. Discussion

The main result of this experiment is the finding that as much as 13% of the variability in performance on the identification task was accounted for by age of the listener, even after taking into account differences in the ability to detect increments in intensity. The goal of Experiment One was to create a situation in which listeners were forced to identify the intensities of two bursts presented sequentially. There was very little energetic overlap between the bursts, and there was no uncertainty about either the order in which the bursts would be presented or about the levels of the baseline and incremented intensities. Even the precise frequency content of the bursts was fixed by using the same three frozen noise tokens throughout the experiment. The IT values are similar to those reported by other researchers (Heinz and Formby, 1999; Cousineau *et al.*, 2009). Nonetheless, many of the listeners who could detect intensity increments of 5–8 dB when the bursts were presented in isolation had substantial difficulty distinguishing among the four possible intensity combinations in the multiburst sequence, for which all of the increments were 15 dB. These results are very similar to the results of Gallun (2010) in terms of the reduction in intensity discrimination ability for even short sequences with long inter-stimulus intervals. There is one main difference, however, which is that in this experiment age impacted the results.

In addition to the change from a comparison between two sequences to the identification of a single sequence, another design difference between the studies involved the number of participants and the greater (although still not complete) dissociation of age and hearing loss. While Gallun (2010) used only 12 participants (five younger and seven

older), the current study substantially improved the power of the analysis by increasing the number of listeners tested to 42. This increased power was important due to the large variability in increment sensitivity in both the single burst and sequential conditions for both the younger and older listeners. By recruiting listeners with more similar hearing ability (both older listeners with better hearing and younger listeners with poorer hearing), it was possible to reduce the degree to which hearing ability was confounded with age as a factor in performance. By using a fixed increment level in the identification task, however, the relationship between increment ability in the single and sequential conditions was preserved, and thus IT did predict the greatest proportion of the variance. The correlation between IT and performance was negative in all cases, suggesting that lower IT values resulted in better performance, as would be expected.

In Gallun (2010), the average performance levels observed were consistent with an ideal-observer model supplied only with the overall level of the stimuli; this implies that some listeners may have been ignoring the sequential nature of the stimulus all together. To examine whether or not this strategy can explain why the older listeners did more poorly than the younger listeners in Experiment One but not in Gallun (2010), it is informative to examine the relative performance for the four different stimuli to be identified. If listeners were simply judging the overall level, they would have been most successful in identifying the no change and both stimuli, which had the highest and lowest combined levels, and would have done most poorly at identifying the middle and high stimuli, which were roughly equal in level and midway between the other no change and both stimuli. Figure 3 and Table I show that the both stimulus was actually the least accurately identified; this is consistent with a strategy of focusing attention on only one of the two bursts. Examination of the confusions among stimuli, shown in Table I, also reveals that, especially for the older listeners, there was a greater likelihood of confusing the both stimulus for the middle or high than for confusing the middle or high with each other. This pattern is consistent with a strategy in which the intensities of the various bursts are identified individually and is inconsistent with a strategy involving the use of overall level.

Taken together, the data of Gallun (2010) and Experiment One are consistent with the results of Trainor and Trehub (1989) as well as Alain *et al.* (1996) in that older listeners were more impaired when the comparisons to be made relied heavily upon auditory memory.

III. EXPERIMENT TWO: DISCRIMINATION OF PRE- AND POST-CUED TARGETS

An important question that is not answered by simply comparing the results of Experiment One with those of Gallun (2010) is whether the difficulty that older listeners experienced was the result of a reduction in the ability to *select* individual elements from a sequence or in the ability to *recall* multiple intensities. The results of Alain *et al.* (1996) would suggest that the selection process is less likely to be impaired than is the processing of stimuli in auditory mem-

ory. To examine this question more carefully, four burst sequences were designed in which the first and fourth bursts were always the low frequency and never targets, while the second and third bursts could be either the middle or high frequency and one of the two was always the target. To indicate which burst was the target, a cue burst was presented either before (“pre-cue”) or after (“post-cue”) the sequence. Each burst was presented at either a high or low level, and listeners were asked to determine whether the cue and the target were the same or different in intensity. The cue was informative of the target frequency, but the temporal position and frequency of the target burst were both fixed across a block of 50 trials. All that changed from trial to trial were the intensities of the various bursts, which were not indicated by the cue.

The key difference between the experiments presented here is that in Experiment One, it was essential that listeners extract the intensities of both the high and middle frequency bursts (essentially a sequential dual-task situation), whereas in Experiment Two, the optimal strategy is to inhibit the processing of all stimuli except the target (a selection task). If the difficulty that older listeners have is in the extraction of a single element from a sequence, then the performance of older listeners should be impaired relative to that of the younger listeners on both the pre- and post-cue conditions. This would be surprising, given the similar performance of older and younger listeners on the sequences used in Gallun (2010). The results of Trainor and Trehub (1989) and of Alain *et al.* (1996) also predict that older listeners are unlikely to be impaired for the sensory processing of sequences. Those studies suggest that it is more likely that aging effects will be found in the later processing stages related to working memory and response selection. A pre-cue was predicted to enhance the saliency of the target and thus aid the listeners in the selection and encoding of a single element. This can be thought of as an “exogenous” cue, which drives attentional selection in a stimulus-based manner by increasing the perceptual saliency of the cued frequency (Desimone and Duncan, 1995; Shinn-Cunningham, 2008). A post-cue, on the other hand, can be thought of as requiring “endogenous” selection based on an internal listener-directed process. While there is little known about the effects of aging on these processes, it would be consistent with the literature to find that an internally driven selection process was more impacted by aging than was a stimulus-driven process.

An important difference between this study and other studies in the literature using cue stimuli (e.g., Hafter *et al.*, 1993; Richards and Neff, 2004) was that the cue, while informative, was redundant with expectations about which burst in the sequence should be extracted and compared in intensity with the cue. Consequently, it is likely that that pre-cue will differ from the post-cue primarily in that the pre-cue will induce involuntary stimulus-based cueing while the post-cue will not. It is hypothesized that the effects of the exogenous pre-cue will be similar for the older and younger listeners but that the endogenous expectation-based cueing will be more effective for the younger listeners, leading to a greater age difference in the post-cue condition.

Another question left unanswered by Experiment One is why older listeners did more poorly when presented with the middle and both sequences as compared with high and no change. changes in the intensity of the middle burst were more difficult to detect for the older listeners, but this could be due to the frequency of the burst, its temporal position, or both. The use of a series of blocked conditions in which frequency and temporal position were both varied was anticipated to reveal whether or not frequency and temporal position had similar impacts on performance in a selection task as in a dual-task where two stimuli must both be processed.

A. Participants

Forty listeners participated, with an average age of 49.3 yr, ranging from 19 to 73 yr. Of these, 20 were younger than 50 yr old (15 males and 5 females) and 20 were 50 yr or older (16 males and 4 females). Hearing sensitivity was no worse than mildly impaired for all participants: average threshold (octave frequencies from 0.25 to 8 kHz) was 15.2 dB HL and ranged from 3.3 to 32.1 dB HL. In this case, the older listeners did not have significantly different average thresholds values from the younger listeners ($t_{40} = 2.00$, $p = 0.06$), but the average threshold for the older listeners (17.1 dB HL) was slightly higher than for the younger (13.3 dB HL). As for the sample of listeners who participated in the first experiment, the greatest differences across groups were in the higher frequencies (4 and 8 kHz), where the ears of the older listeners were on average 5–19 dB less sensitive than were those of the younger listeners. For most impaired listeners, losses were symmetrical across ears (no differences greater than 20 dB at any frequency), but three impaired listeners had relatively normal hearing in either the right or left ear. All listeners were paid for their participation and all procedures were reviewed and approved by the Institutional Review Boards of both the Portland VA Medical Center and the Oregon Health and Science University.

B. Stimuli

Stimuli were 50-ms noise bursts with 5 ms $\cos^2$ onset and offset ramps. Bursts were presented in three frequency ranges: low (0.40–0.56 kHz), middle (1.62–2.40 kHz), or high (4.08–6.10 kHz). One possibility for the greater effect of age in Experiment One relative to the data of Gallun (2010) was the use of frozen noise. Consequently, this factor was varied, and on each trial, listeners either heard a combination of the same three “fixed” noise bursts used in Experiment One or new “random” bursts generated for that trial. All noise bursts were generated as in Experiment One and presented using the equipment described in the preceding text.

C. Procedures

Experiment Two was divided into the same three parts as was the first experiment: detection of single bursts, discrimination of single-burst intensities, and discrimination of the intensities of bursts presented as sequences. DTs and ITs

were obtained using a similar adaptive tracking procedure as in the first experiment. As the identification task involved longer testing, a two-down/one-up adaptive tracking procedure was used instead of three-down/one-up, which leads to an estimate of the level that would result in 70.1% correct performance rather than 79.4% (Levitt, 1971). DTs and ITs were calculated on the basis of four adaptive tracking blocks for each noise burst frequency, for a total of 12 blocks of each. Fixed noise bursts were presented on half of the blocks and random bursts on the other half, but no significant differences were observed ($p > 0.05$), and so the thresholds were combined in order to provide more accurate estimates of performance. Only the combined thresholds are reported in the following text.

The third part of the experiment was a fixed-increment identification task in which participants compared the intensity of a single noise burst (the “cue”) with that of a “target,” which could be either the second or third noise burst in a sequence of four noise bursts. One of the middle two bursts was the mid frequency burst and the other was the high frequency burst. To ensure that listeners processed the intensity of the cue and the target, either, both, or neither could contain increments. The non-target middle or high frequency burst was randomly incremented or unincremented, providing a distracting context for the target burst. The low frequency burst that started and ended the sequence was always presented at the standard level of 30 dB SL.

To ensure that increments were equally detectable, regardless of baseline IT, each listener’s stimuli were created individually and increments were set to the IT for that listener for that frequency plus another 10 dB. For example, a listener with a single-burst IT of 4 dB for the middle frequency would have heard middle-frequency increments of 14 dB in the sequential condition, while a listener with an IT of 8 dB would have heard increments of 18 dB. An alternative approach, which was used in the first experiment, would have been to use the same size of increment for every listener. Because sensitivity to increments was such a strong predictor of performance, it was hoped that this factor could be mitigated by providing each listener with an increment that was known to be substantially greater than what that listener could detect. As the results in the following text will show, however, even when listeners with lower intensity discrimination thresholds were given smaller increments, they outperformed listeners who had higher thresholds and were given larger increments.

Each trial consisted of two intervals with a 500-ms inter-stimulus interval (ISI) between them. One interval contained the 50-ms cue followed by 450 ms of silence, creating a 500-ms “cue interval” indicated by a visual marker. The other interval contained a sequence of four 50-ms bursts separated by 150 ms of silence, resulting in a 650-ms “target interval,” also indicated by a visual marker. In the pre-cue condition, the cue interval preceded the ISI, whereas in the post-cue condition, the target interval preceded the ISI. This resulted in a total trial time of 1650 ms in both conditions. A schematic is provided in Fig. 4, where two “same” trials are shown in the top panels and two “different” trials in the bottom panels. Shading of the elements indicates frequency,

FIG. 4. Schematic description of the stimuli used in the same/different identification task in Experiment Two. Each element corresponds to a burst of noise, with frequency indicated by shading and intensity indicated by size, as in Fig. 1. “Same” trials (top panels) are defined by the same sizes (intensities) of the elements in the cue and stimulus boxes (marked by lights in the experiment). “Different” trials (bottom panels) are defined by the fact that the cue and stimulus elements that match in frequency do not match in intensity.

and size indicates intensity. On the “same” trials, the element with the same shading (frequency) as the element in the “cue” interval (which was marked by a light) has the same size (intensity) as the cue, indicating that the listener would be correct if they responded “same.” On the “different” trials, the element of the stimulus that matches the frequency of the cue element is a different intensity, indicating that the listener should respond “different.” The first and fourth bursts were always low frequency, and the task of the listener was to judge the intensity of either the second or third burst relative to the intensity of the cue. The frequencies of the second and third bursts were varied across blocks of trials with one being a middle-frequency burst and the other a high-frequency burst. For each block of trials, a cue was either presented before the sequence began or after it had ended, and the frequency of the cue always matched that of either the second or third burst. As mentioned in the preceding text, pilot testing suggested that varying the frequency of the target and its position in the sequence across trials was too difficult. For this reason, blocks of 50 trials were presented in which the frequency of the target (middle or high) and its position in the sequence (which will be referred to for simplicity as the “first” or “second” position) were fixed across blocks. In terms of information, then, there was no uncertainty as to the target frequency or position after the first trial in a block.

To respond correctly, the listener had to report whether the burst with a center frequency matching that of the cue had also been presented at a level matching that of the cue. When the cue and target levels matched in the “same” condition, they could both be the standard level (30 dB SL) or they could both be incremented. When they did not match, the “different” condition, one was presented at the standard level and the other was incremented. On half of the blocks the noise bursts were generated randomly on each trial, whereas on the other blocks the same set of frozen noises were used for the entire block.

As can be seen by the schematic diagrams of the trial timing shown in Fig. 4, the requirement that the cue consist of a 500-ms marked interval in both pre- and post-trial cueing conditions introduced a substantial difference in timing between the cue and target in the two conditions. Pilot data were collected with shorter cue intervals and with cue elements that did not share onsets with the marker to remove this difference. Subjects who already found the task very challenging were confused by the uncertain occurrence of the cue in the middle of the marked interval and found a 50-ms cue interval to be too brief to adequately perform the task. With more training, these aspects could have been eliminated, but the decision was made to remove any design aspects that might have increased confusion among the older listeners while sparing the younger. An analysis of the

data with regard to timing differences (see Sec. III D) suggests that the time between cue and target had no reliable effect on performance.

Across the 32 experimental blocks, each of which contained 40 trials, four factors were each varied across two possible values (for 16 possible combinations), and each combination was tested twice. The factors (and possible values) were: cue (pre or post), center frequency (middle or high), position (first or second), and noise (fixed or random). Each block contained only a single combination of the two possible values associated with each of the four factors and each combination was tested in two blocks. The order of the blocks was randomized anew for each subject. Testing occurred over two to three 2-h sessions.

D. Results

As in Experiment One, detection thresholds were fairly low for all three bursts and differed between the older and younger groups. A mixed-models ANOVA with age group as a between-subjects factor and frequency as a repeated-measures factor found a significant effect of frequency ($F_{2,76} = 5.92$, $p = 0.004$), a significant effect of age group ($F_{1,38} = 4.54$, $p = 0.04$), and no significant interaction ($F_{2,76} = 1.40$, $p = 0.25$). For the younger group, DTs were 6.4, 7.5, and 8.6 dB for the low, middle, and high bursts, respectively, whereas for the older group, DTs were 8.3, 9.9, and 14.1 dB. Average DTs for the two groups are plotted in Fig. 5(a). As age and pure-tone thresholds were significantly correlated, especially for the higher frequencies, stepwise multiple linear regression was again used to determine whether the main effect of age was due to differences in pure-tone thresholds or whether age of the listener was truly associated with performance. As in the first experiment, pure-tone thresholds were strong predictors and age explained no significantly greater proportion of the variance. The predictors were again consistent with the frequency content of the bursts. DT for the low burst was best predicted by pure-tone thresholds at 500 and 1000 Hz ($R^2 = 0.64$, $p < 0.001$), DT for the mid burst was best predicted by pure-

tone thresholds at 1000 and 2000 Hz ($R^2 = 0.74$, $p < 0.001$), and DT for the High burst was best predicted by pure-tone thresholds at 4000 and 8000 Hz ($R^2 = 0.71$, $p < 0.001$). Additional predictors did not significantly improve the predictions of the model.

Increment thresholds for the three bursts differed with frequency but were fairly similar for the older and younger listeners. A mixed-models ANOVA with age group as a between-subjects factor and frequency as a repeated-measures factor found a significant effect of frequency ($F_{2,80} = 33.93$, $p < 0.001$), no significant effect of age group ($F_{1,40} = 0.07$, $p = 0.79$), and no significant interaction ($F_{2,80} = 0.90$, $p = 0.41$). For the younger listeners, ITs were 4.5, 6.0, and 6.1 dB for the low, middle, and high bursts, respectively, while for the older listeners the same bursts produced thresholds of 4.5, 6.1, and 5.8 dB. Average ITs for the two groups are plotted in Fig. 5(b).

To examine performance in the cueing task with sequentially presented bursts, performance on the same/different task was converted to d' by combining selected subsets of the trials. Trials on which the cue and target were different in intensity (regardless of whether the cue or the target was the more intense) were treated as one type of trial ("signal" in the terminology of signal detection theory), and trials on which the cue and target were the same were treated as the other type ("noise" trials). In this manner, each listener response was classified as a hit, miss, correction rejection, or false alarm and the d' metric was calculated for each combination of cue, center frequency, target position, and noise type for the older and younger listeners. Those values are shown in Table II.

A mixed-models ANOVA was conducted with age group as a between-subjects factor and cue, center frequency, target position, and noise type as repeated-measures factors. This statistical approach revealed significant interactions between cue and center frequency ($F_{1,38} = 11.40$, $p = 0.002$) as well as between cue and age group ($F_{1,38} = 6.19$, $p = 0.02$). Main effects of cue ($F_{1,38} = 46.34$, $p < 0.001$), center frequency ($F_{1,38} = 6.57$, $p = 0.01$), and position ($F_{1,38} = 5.80$, $p = 0.02$) were also revealed. The effect of noise type was not significant ($p = 0.93$) nor was the main effect of age group ($p = 0.25$). Figure 6 shows the average performance values for the two age groups as a function of whether the cue was presented before or after the target sequence, for the two center frequencies, and as a function of target position. Average d' values for the full set of parameters are shown in Table II.

To determine the influence of age on performance, the data were averaged across frequency, position, and noise type, allowing an analysis of pre- and post-cue performance for the individual listeners. Multiple linear regression was used to predict the individual d' values for the pre- and post-cue conditions, with DTs, ITs, and current age of the listener entered as potential predictors. The pre-cue performance was not significantly correlated with age ($r = -0.02$, $p = 0.45$), and the majority of the variance was accounted for by IT for the high burst (45%), with DT for the middle burst accounting for an additional 8% ($R^2 = 0.53$, $p < 0.001$). None of the other predictors significantly improved the model. The post-cue performance was significantly correlated with age

FIG. 5. Detection thresholds (a) and increment thresholds (b) for older (gray circles) and younger listeners (black squares) in Experiment Two. Error bars indicate two standard errors of the mean.

TABLE II. Sensitivity (d') to the relationship between the intensity of the cue and the target as a function of cue location (pre-trial vs post-trial), frequency (middle vs high), position in the sequence (first vs second) and noise type (random vs fixed) for older and younger listeners. Mean performance for all listeners and 1 standard error of the mean (SE) are shown in the final two columns.

Cue	Frequency	Target Position	Noise	Older	Younger	Mean	SE
Pre	Middle	First	Random	2.19	2.41	2.28	0.20
			Fixed	2.09	2.42		
		Second	Random	1.96	2.17	2.06	0.20
			Fixed	2.06	2.07		
	High	First	Random	2.22	2.50	2.41	0.20
			Fixed	2.39	2.51		
		Second	Random	2.04	2.08	2.13	0.22
			Fixed	2.20	2.18		
Post	Middle	First	Random	1.11	1.73	1.33	0.15
			Fixed	0.87	1.62		
		Second	Random	0.80	1.58	1.17	0.16
			Fixed	0.75	1.56		
	High	First	Random	1.52	1.99	1.78	0.20
			Fixed	1.67	1.95		
		Second	Random	1.44	2.03	1.70	0.20
			Fixed	1.29	2.04		

($r = -0.29$, $p = 0.035$) although as in the first experiment, the majority of the variance was accounted for by IT for the high burst (31%) with age accounting for 16% of the remaining variance ($R^2 = 0.47$, $p = 0.002$). None of the other predictors significantly improved the model. The correlations between IT and performance were negative for all three IT values ($p < 0.05$ in all cases), indicating that even though lis-

teners with lower ITs were given smaller increments, they outperformed listeners who had higher ITs and were given larger increments.

As mentioned in the preceding text, one potentially important difference between the pre- and post-cue conditions was the total time between cue and target. Because the cue interval was 500 ms, but the cue itself only 50 ms, a difference existed in the time that elapsed between the target and cue depending on whether the cue preceded or followed the target (see Fig. 4). In the pre-cue condition, the time between the offset of the cue and the onset of target sequence was 950 ms, while in the post-cue condition, the time from the offset of the target sequence to the onset of the cue was 500 ms. In the pre-cue condition, this made the time between the offset of the cue and the onset of the second burst 1150 ms and the time to the onset of the third burst 1350 ms. In the post-cue condition, however, the time between the offset of the second burst and the onset of the was 900 ms and the time from the offset of the third burst to the cue was 700 ms. If the delay between target and cue was the determinant of the differences in performance in the pre- and post-cue conditions, this would also predict that listeners would have performed better with longer delays and worse with shorter.

To examine the effect of delay time on performance, the data were averaged across frequency and noise type, allowing a comparison of cue and position, with age as a potential between-subjects factor. Average d' values for the pre-cue in the first and second positions were 2.3 and 2.1, while values for the post-cue in those positions were 1.6 and 1.4. Statistically, there was no interaction between cue and position ($F_{1,38} = 1.47$, $p = 0.233$) nor among cue, position, and age ($F_{1,38} = 2.59$, $p = 0.116$), although the main effects of cue and position remained significant, as well as the interactions between age and cue observed in the full data set. These results are not consistent with the hypothesis that performance in the pre-trial cue condition was better than with the post-trial cue because a longer time between cue and target led to better performance. Rather, they suggest that regardless of cue position, the target was better identified when it occurred as the second of the four bursts in the sequence.

E. Discussion

There were two goals of this experiment. The primary goal of Experiment Two was to better understand the effect of age on the selective processing of auditory sequences as opposed to the dual-task condition of Experiment One. Figure 6 reveals that while the older listeners performed more poorly on all of the conditions, the effect was greatest in the post-cue conditions. For younger listeners, the difference in d' between the pre-cue and post-cue conditions was significant (2.3 vs 1.8), but for the older listeners the difference was both significant and doubled in size (2.2 vs 1.2).

While it is clearly the case that the post-cue condition was more difficult, which may in itself explain much of the difference between younger and older listeners, it would be useful to identify more clearly exactly what feature of the post-cue condition was responsible for the difficulties

FIG. 6. Accuracy in the same/different judgments required in Experiment 2, averaged across conditions. The first panel shows the effect of cue, averaged across noise type (random vs frozen), frequency, and position. The second and third panels show the same data averaged to reveal the effects of frequency and position, respectively. Older listeners are indicated by gray circles and younger by black squares. The significant interactions of cue and age group are illustrated by the small difference between the groups in the pre-trial cue condition and the larger difference in the post-trial cue condition. Error bars indicate two standard errors of the mean.

listeners experienced. In many experiments, a cue is used to reveal which of the stimulus elements was the target. In Experiment Two, however, the stimulus conditions were presented in blocks, thus greatly reducing listener uncertainty about the target frequency and position. If the target was the middle frequency in the second position on the first trial, then the middle frequency in the second position would be the target for the remainder of the 50 trials in that block. This suggests that it was not primarily uncertainty about the position or frequency of the target that caused the listeners to have greater difficulty in the post-cue than in the pre-cue conditions.

Due to the fixed nature of the design, it cannot be the case that the pre-cue was fundamentally more informative than the post-cue. Indeed there was very little information to be gained from the cue. Rather it is likely that the sensory memory of the pre-cue is what is useful. Given this situation, it is worth considering the possibility that the post-cue on a given trial actually was functioning as a pre-cue for the next trial. This is only possible because the cue frequency and position (post-cue or pre-cue) was fixed across trials in a block as was the position of the target in the sequence. That performance was worse in the post-cue than in the pre-cue conditions suggests that the post-cue cannot serve as an effective pre-cue for the next trial or that the pre-cue effect is limited in duration. If the duration of the pre-cue effect is limited, then it is possible that the few seconds it takes for the subject to respond and receive feedback degrades the sensory memory for the cue to the point where it no longer enhances performance. Alternatively, it is possible that the post-cue is an effective pre-cue for the stimulus but that the effect is reduced, and the difference between pre-cue and post-cue would have been greater if a longer intertrial interval (or a between-trial masker) had been used. Future work examining these issues is clearly warranted.

These are not the only studies that have examined the impact of age on the processing of stimulus sequences. Trainor and Trehub (1989) and Alain *et al.* (1996) both examined the effects of aging on the ability to select and process auditory targets by using two interleaved sequences and asking listeners to make judgments about the elements of one of the sequences. The results showed that larger frequency differences between the two sequences led to better target processing for all listeners but that regardless of frequency difference, the older listeners performed more poorly than did the younger. This was interpreted as evidence that the initial stimulus-based sequence processing was not impaired by aging but that the later processing of the target elements was less accurate for the older listeners. The stimulus-based streaming of the sequence elements can be considered to be similar to the pre-cue, which resulted in similar performance across older and younger listeners, while the requirement that the listeners use an internal expectation to select the target from the sequence in the post-cue may be considered to be similar to the processing of individual elements in the earlier studies. Viewed in this manner, the data presented here are consistent with those in the literature suggesting that aging has its impact by reduc-

ing the effectiveness or efficiency of internal processes such as auditory memory and/or endogenous attentional selection.

One possible framework for understanding the age-related declines in performance is what McKeown and colleagues (McKeown and Wellsted, 2009; Mercer and McKeown, 2010) call the timbre memory model. According to this framework, memory for the intensity profile of recently presented stimuli persists for roughly 20 s and is influenced both by the history of sounds that have been presented in this time period and by internally directed attention to these persistent sensory memories. Such a framework can be used to explain both the enhancement of stimuli through attention as well as the interference through automatic updating of the representations, which can overwrite the sensory memory of older sounds. The current results, seen through the lens of this model, could be taken as evidence that when a sensory memory of the target is created by a pre-cue, both older and younger listeners can effectively attend to that frequency and subsequently make the cue-target comparison, but when no sensory memory exists, older listeners are less able to generate an internal enhancement of the target and thus are less effective at making the subsequent comparison.

A secondary goal of Experiment Two was to more carefully examine the relative importance of temporal position and center frequency in the sequence identification results of Experiment One. As can be seen in Fig. 6, averaging across all of the other factors reveals that both center frequency and temporal position are important and that older listeners were not fundamentally different from younger listeners in terms of which conditions were the most difficult. Center frequency followed the pattern shown in Experiment One, with better performance on high frequency bursts than on middle frequency bursts. For temporal position, however, the best performance was for the first target rather than the second target. This result, which is the reverse of what was observed in Experiment One, suggests that adding the final low tone to the sequence probably reduced performance for the second target. There was no significant interaction between the factors. These results suggest that when both temporal position and center frequency are manipulated, each can independently influence the ability of listeners to process the intensity of the elements in a sequence. Furthermore, the only significant difference between the age groups was found with the cueing manipulation.

IV. SUMMARY

The goal of these experiments was to determine the extent to which aging leads to reduced performance in intensity discrimination for target elements embedded in a rapidly presented auditory sequence. The data are consistent with the idea that the internal selection and/or processing of auditory stimuli is more difficult for older than for younger listeners. While intensity discrimination sensitivity was not different for the two age groups, identification of sequences was influenced by listener age. To control for potential deficits in the ability to extract information from rapid sequences, a cueing paradigm was developed that used identical target sequences to produce an age-related deficit when a

post-cue was employed but not when the task was conducted using a pre-cue. Further experiments are necessary to more fully describe the nature of the interference, the stimuli for which such interference is observed, and the relationship between interference on psychophysical tasks and more naturalistic tasks such as comprehension of speech. It seems likely, however, that deficits such as those uncovered here may provide insight into speech understanding difficulties experienced by older listeners with relatively intact pure tone sensitivity.

ACKNOWLEDGMENTS

This work was supported by the Department of Veterans Affairs, Veterans Health Administration, Rehabilitation Research and Development Service (VA RR&D Career Development II Award C4963W to F.J.G.) and by NIH/NIDCD (R03 DC08395 to F.J.G.). Facilities and equipment support was also provided by the VA RR&D National Center for Rehabilitative Auditory Research. The authors are extremely grateful to Marjorie Leek for help with the sensitivity analysis, to Stephen Fausti for guidance, to Matthew Marble, Kelly Reavis, Marc Caldwell, and Erin Connor for help with data collection, to Marilyn Dille and Julia Fitzer for discussions of design issues, and of course to our listeners. Two anonymous reviewers gave very helpful comments on previous versions of the manuscript and one of the anonymous reviewers provided an especially useful analysis of the [Eguchi and Hirsh \(1969\)](#) method of calculating listener sensitivity, the core insights of which are described in the Appendix. The opinions and assertions presented are private views of the authors and are not to be construed as official or as necessarily reflecting the views of the Department of Veterans Affairs.

APPENDIX

Converting between percent correct and d' for a four-alternative task presents a number of logic difficulties (see [Hacker and Ratcliff, 1979](#); [Green and Dai, 1991](#); [Macmillan and Creelman, 2005](#)). The data from Experiment One are presented in the preceding text both as percent correct values (Table I) and as d' values, based on the method of [Eguchi and Hirsh \(1969\)](#), which was brought to our attention by [Leek et al. \(1987\)](#). The d' analysis involved calculating two values for each stimulus with the first value (the number of hits) corresponding to the number of trials on which a given response was emitted when that stimulus had actually been presented. The second value (the number of false alarms) corresponded to the number of trials on which that same response was emitted, but any one of the other three stimuli had been presented. The hits and false alarms rates were calculated by dividing the number of hits by the total number of trials on which the given stimulus was actually presented and dividing the number of false alarms by the total number of trials on which any of the other stimuli were presented. The rates were then converted into z scores and the difference between those z scores was used as a measure of the discriminability of each stimulus relative to the other three stimuli. The use of hits and false alarms to calculate d'

avoids the problem of using the same trials for multiple calculations, because both probabilities are based only on the trials on which a given stimulus value was given as the response.

The Eguchi and Hirsh approach reflects the accuracy with which participants were able to associate a given label with the correct stimulus regardless of the number of times that label was chosen and thus is an improvement over the percent correct measure, which is just the “hit rate” and is thus influenced by the number of times a given response is emitted. There is a drawback to the Eguchi and Hirsh approach, however, which is that if an observer is biased to report one stimulus more often than the others, then the relative proportion of trials on which that stimulus is presented will influence the false alarm rate. In other words, if an observer is more likely to report a given stimulus, then the chance of a false alarm will decrease as the likelihood of that stimulus having been presented increases.

Bias is clearly present in Experiment One, as is shown in Table I, because the proportion of the trials on which observers labeled the change stimuli (middle, high, or both) as the no change stimulus (up to 0.38 for older subjects with the middle stimulus) is substantially greater than the proportion of trials on which no change stimulus was labeled as any one of the change stimuli (0.05 for older subjects with the middle stimulus). Because the no change stimulus was presented on 55% of the trials, the total number of errors was lower than if, for example, the middle stimulus had been presented on 55% of the trials. The Eguchi and Hirsh calculation does not compensate for this factor.

To examine the impacts of the Eguchi and Hirsh calculation method on the data presented in Experiment One, two analyses were conducted. In the first, percent correct values from Table I were transformed into d' values using the Hacker and Ratcliff calculation, based on their published table. In the second, two simulations were conducted by applying the proportions shown in Table I to two sets of stimulus distributions and then using the Eguchi and Hirsh calculation method to examine the effects of stimulus distribution on the resulting values. The first simulation used a set of equally distributed stimuli and the second simulation used a set of stimuli distributed asymmetrically in the manner used in Experiment One (55% of the trials were no change stimuli and 15% were each of the other three stimulus types).

The resulting values showed the same basic patterns as appear in the percent correct data [Fig. 3(a)], as well as in the data transformed to d' values using the Eguchi and Hirsh approach and then averaged across subjects [Fig. 3(b)]. For all of the change stimuli, the distributions used in Experiment One resulted in d' values that were *lower* than were obtained for equally distributed stimuli by 0.06–0.22 and *higher* than the values obtained in the Hacker and Ratcliff table by 0.04–0.47. For the no change stimulus, there was no difference between the values calculated for the two sets of distributions; this was due to the fact that the differences in the two simulations were driven entirely by the number of no change trials included in the false alarm calculations. In comparison with the Hacker and Ratcliff tables, however, the Eguchi and Hirsh approach resulted in d' values that

were *lower* by 0.25 for the younger listeners and by 0.31 for the older listeners. Although this suggests that the bias in the data impacted the resulting d' values, it is not obvious that the Hacker and Ratcliff approach, which assumes no bias, would have provided a more accurate measure of sensitivity. From these simulations, it appears that the Hacker and Ratcliff approach systematically underestimates sensitivity in cases where biased responding increases false alarms (the change stimuli) and overestimates sensitivity in cases where biased responding decreases false alarms (the no change stimuli).

Most importantly for the main aim of this study, however, the differences in sensitivity between the older and younger listeners were only minimally affected by the stimulus distribution used in Experiment One but would have been underestimated by sensitivity calculations based on the Hacker and Ratcliff method. The simulation based on the set of distributions employed in Experiment One predicted d' values that were on average 0.79 greater for the younger listeners. The simulation based on equal distributions of stimuli predicted d' values that were on average 0.74 greater for the younger listeners. The Hacker and Ratcliff calculation predicted d' values that were on average 0.61 greater for the younger listeners. The data shown in Fig. 3(b) demonstrate an average difference of 0.81 between older and younger listeners. From this we conclude that the differences between calculation methods, while worth noting, are substantially smaller than the effect sizes with which this report is concerned.

Alain, C., Ogawa, K. H., and Woods, D. L. (1996). "Aging and the segregation of auditory stimulus sequences," *J. Gerontol. Psychol. Sci.* **51**, 91–93.

Chi, T., Gao, Y., Guyton, M. C., Ru, P., and Shamma, S. (1999). "Spectrotemporal modulation transfer functions and speech intelligibility," *J. Acoust. Soc. Am.* **106**, 2719–2732.

Cousineau, M., L., Demany, L., and Pressnitzer, D. (2009). "What makes a melody: The perceptual singularity of pitch sequences," *J. Acoust. Soc. Am.* **126**, 3179–3187.

Cowan, N. (1984). "On short and long auditory stores," *Psychol. Bull.* **96**, 341–370.

Cowan, N. (1988). "Evolving conceptions of memory storage, selective attention and their mutual constraints within the human information processing system," *Psychol. Bull.* **104**, 163–191.

Cowan, N. (1995). *Attention and Memory: An Integrated Framework* (Oxford University Press, New York).

Demany, L., and Semal, C. (2008). "The role of memory in auditory perception," in *Auditory Perception of Sound Sources*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer Science and Business Media, New York), pp. 77–113.

Desimone, R., and Duncan, J. (1995). "Neural mechanisms of selective visual attention," *Annu. Rev. Neurosci.* **18**, 193–222.

Eguchi, S., and Hirsh, I. J. (1969). "Development of speech sounds in children," *Acta Oto-Laryngol., Suppl.* **257**, 1–51.

Espinoza-Varas, B., and Watson, C. S. (1986). "Temporal discrimination for single components of nonspeech auditory patterns," *J. Acoust. Soc. Am.* **80**, 1685–1694.

Fitzgibbons, P. J., and Gordon-Salant, S. (1996). "Auditory temporal processing in elderly listeners," *J. Am. Acad. Audiol.* **7**, 183–189.

Fitzgibbons, P. J., and Gordon-Salant, S. (1998). "Auditory temporal order perception in younger and older adults," *J. Speech Lang. Hear. Res.* **41**, 1052–1060.

Fitzgibbons, P. J., and Gordon-Salant, S. (2001). "Aging and temporal discrimination in auditory sequences," *J. Acoust. Soc. Am.* **109**, 2955–2963.

Fitzgibbons, P. J., Gordon-Salant, S., and Barrett, J. (2007). "Age-related differences in discrimination of an interval separating onsets of successive tone bursts as a function of interval duration," *J. Acoust. Soc. Am.* **122**, 458–466.

Gallun, F. J. (2010). "Judgments of intensity for brief sequences," *J. Acoust. Soc. Am.* **127**, EL166–171.

Gordon-Salant, S., Yeni-Komishian, G. H., Fitzgibbons, P. J., and Barrett, J. (2006). "Age-related differences in identification and discrimination of temporal cues in speech segments," *J. Acoust. Soc. Am.* **119**, 2455–2466.

Green, D., and Dai, H. (1991). "Probability of being correct with 1 of M orthogonal signals," *Percept. Psychophys.* **49**, 100–101.

Hacker, M. J., and Ratcliff, R. (1979). "A revised table of d' for M-alternative forced choice," *Percept. Psychophys.* **26**, 168–170.

Haft, E. R., Schlauch, R. S., and Tang, J. (1993). "Attending to auditory filters that were not stimulated directly," *J. Acoust. Soc. Am.* **94**, 743–747.

Heinz, M. G., and Formby, C. (1999). "Detection of time- and bandlimited increments and decrements in a random-level noise," *J. Acoust. Soc. Am.* **106**, 313–326.

Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (2008). "Informational masking," in *Auditory Perception of Sound Sources*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer Science and Business Media, New York), pp. 143–189.

Leek, M. R., and Watson, C. S. (1984). "Learning to detect auditory pattern components," *J. Acoust. Soc. Am.* **76**, 1037–1044.

Leek, M. R., Dorman, M. F., and Summerfield, Q. (1987). "Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **81**, 148–154.

Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.

Macmillan, N. A., and Creelman, C. D. (2005). *Detection Theory: A User's Guide* (Lawrence Erlbaum Associates, New York), Chap. 10, pp. 245–255.

McKeown, D., and Wellsted, D. (2009). "Auditory memory for timbre," *J. Exp. Psych: Hum. Percept. Perform.* **35**, 855–875.

Mercer, T., and McKeown, D. (2010). "Updating and feature overwriting in short-term memory for timbre," *Atten. Percept. Psychophys.* **72**, 2289–2303.

Richards, V. M., and Neff, D. L. (2004). "Cueing effects for informational masking," *J. Acoust. Soc. Am.* **115**, 289–300.

Salthouse, T. A. (2000). "Aging and measures of processing speed," *Biol. Psychol.* **54**, 35–54.

Shinn-Cunningham, B. G. (2008). "Object-based auditory and visual attention," *Trends Cogn. Sci.* **12**, 182–186.

Spiegel, M. F., and Watson, C. S. (1981). "Factors in the discrimination of tonal patterns. III. Frequency discrimination with components of well-learned patterns," *J. Acoust. Soc. Am.* **69**, 223–230.

Trainor, L. J., and Trehub, S. E. (1989). "Aging and auditory temporal sequencing: Ordering the elements of repeating tone patterns," *Percept. Psychophys.* **45**, 417–426.

Tremblay, S., Vachon, F., and Jones, D. M. (2005). "Attentional and perceptual sources of the auditory attentional blink," *Percept. Psychophys.* **67**, 195–208.

Vachon, F., and Tremblay, S. (2005). "Auditory attentional blink: Masking the second target is necessary, delayed masking is sufficient," *Can. J. Exp. Psychol.* **59**, 279–286.

Vachon, F., and Tremblay, S. (2006). "Delayed masking and the auditory attentional blink: A test for retrieval competition and bottleneck models," *Exp. Psychol.* **53**, 182–190.

Vachon, F., and Tremblay, S. (2008). "Modality-specific and amodal sources of interference in the attentional blink," *Percept. Psychophys.* **70**, 1000–1015.

Vachon, F., Tremblay, S., and Jones, D. M. (2007). "Task-set reconfiguration suspends perceptual processing: evidence from semantic priming during the attentional blink," *J. Exp. Psychol. Human.* **33**, 330–347.

Watson, C. S. (2005). "Some comments on informational masking," *Acta Acust. Acust.* **91**, 502–512.

Watson, C. S., Wroton, H. W., Kelly, W. J., and Benbassat, C. A. (1975). "Factors in the discrimination of tonal patterns. I. Component frequency, temporal position, and silent intervals," *J. Acoust. Soc. Am.* **57**, 1175–1185.

Wingfield, A., Tun, P. A., Koh, C. K., and Rosen, M. J. (1999). "Regaining lost time: adult aging and the effect of time restoration on recall of time-compressed speech," *Psychol. Aging* **14**, 380–389.