JARO

Journal of the Association for Research in Otolaryngology

Spatial Selective Auditory Attention in the Presence of Reverberant Energy: Individual Differences in Normal-Hearing Listeners

DOREA RUGGLES AND BARBARA SHINN-CUNNINGHAM

Hearing Research Center, Departments of Biomedical Engineering and Cognitive and Neural Systems, Boston University, 677 Beacon St, Boston, MA 02215, USA

ABSTRACT

Listeners can selectively attend to a desired target by directing attention to known target source features, such as location or pitch. Reverberation, however, reduces the reliability of the cues that allow a target source to be segregated and selected from a sound mixture. Given this, it is likely that reverberant energy interferes with selective auditory attention. Anecdotal reports suggest that the ability to focus spatial auditory attention degrades even with early aging, yet there is little evidence that middle-aged listeners have behavioral deficits on tasks requiring selective auditory attention. The current study was designed to look for individual differences in selective attention ability and to see if any such differences correlate with age. Normal-hearing adults, ranging in age from 18 to 55 years, were asked to report a stream of digits located directly ahead in a simulated rectangular room. Simultaneous, competing masker digit streams were simulated at locations 15° left and right of center. The level of reverberation was varied to alter task difficulty by interfering with localization cues (increasing localization blur). Overall, performance was best in the anechoic condition and worst in the high-reverberation condition. Listeners nearly always reported a digit from one of the three competing streams, showing that reverberation did not render the digits unintelligible. Importantly, inter-subject differences were extremely large. These differences, however, were not significantly correlated with age, memory span, or hearing status. These results show that listeners with audiometrically normal pure tone thresholds differ in their ability to selectively attend to a desired source, a task important in everyday communication. Further work is necessary to determine if these differences arise from differences in peripheral auditory function or in more central function.

Keywords: aging, reverberation, perceptual organization, auditory object, memory

INTRODUCTION

Listening to one talker in a setting where different conversations jockey for attention is one of the most demanding auditory tasks we undertake. Performing this feat depends on selective auditory attention, the process that allows listeners to filter out unwanted sounds and focus on a desired source (Cherry 1953; Bronkhorst 2000; Shinn-Cunningham 2008; Hill and Miller 2010). Selective auditory attention requires listeners both to perceptually segregate a source of interest from competing sources (forming a distinct perceptual auditory object; Darwin and Carlyon 1995) and to bring the desired auditory object into the perceptual foreground (e.g., see Francis 2010; Best et al. 2008; Ihlefeld and Shinn-Cunningham 2008). For instance, individuals can listen for an object with a particular attribute (such as focusing on the source from straight ahead), which helps bring that object into attentional focus (e.g., see Freyman et al. 1999;

Correspondence to: Barbara Shinn-Cunningham · Hearing Research Center, Departments of Biomedical Engineering and Cognitive and Neural Systems · Boston University · 677 Beacon St, Boston, MA 02215, USA. Telephone: +1-617-3535764; fax: +1-617-3537755; email: shinn@cns.bu.edu

Darwin et al. 2003; Kidd et al. 2005a; Marrone et al. 2008a, b; Shinn-Cunningham and Best 2008).

While most listeners are good at directing spatial selective auditory attention, some environments make this task more difficult than others. For instance, reverberant energy distorts the signals reaching the ears (Nabelek et al. 1989; Darwin and Hukin 2000; Lavandier and Culling 2007, 2008). Both spatial cues (such as interaural time differences or interaural level differences) and pitch cues are less reliable in the presence of reverberant energy (Darwin and Hukin 2000; Culling et al. 2003; Zurek et al. 2004). These degradations can hinder object formation and object selection, interfering with selective attention (Darwin and Hukin 2000; Culling et al. 2003; Kidd et al. 2005b; Lavandier and Culling 2008; Marrone et al. 2008a, b).

Aging can also impede spatial selective auditory attention (e.g., Snell et al. 2002; Tun et al. 2002; Helfer and Freyman 2008; Marrone et al. 2008a, b; Singh et al. 2008). Anecdotally, even middle-aged listeners report problems with selective auditory attention (Helfer and Freyman 2008; Wambacq et al. 2009). A recent study shows that interaural phase discrimination is poorer in middle-aged listeners (40–55 years) than in young listeners (18–27 years; it is poorer still in older listeners aged 63–75), suggesting that temporal fine structure processing degrades at an earlier age than is often assumed for other forms of presbycusis (Grose and Mamo 2010). Other recent studies also show that middle-aged listeners perform worse than younger listeners in some tasks, particularly those involving temporal processing (Grose et al. 2009; Helfer and Vargo 2009). Consistent with this, brainstem responses to interaural time differences (ITDs) decline with age, and such deficits have been documented in middle-aged listeners (Ross et al. 2007; Ross 2008; Wambacq et al. 2009). This combination of behavioral and physiological evidence suggests that a task that relies on spatial selective auditory attention may be particularly sensitive in revealing effects of early aging on auditory perception.

This study measured spatial selective auditory attention in listeners ranging in age from young adulthood to early middle age (ages studied in Helfer and Vargo 2009; Grose and Mamo 2010); this choice allowed us to study a adult listeners over a large age span (almost four decades) while avoiding confounding factors such as reduced cognitive function (which is apt to contribute to differences in performance in listeners older than those in our cohort; e.g., see Pichora-Fuller and Singh 2006; Sheldon et al. 2008a, b; Arlinger et al. 2009). We reasoned that any early aging effects on the auditory system were most likely to be revealed in conditions in which fine temporal cues (that are used to compute ITD) are not robustly represented in the acoustic inputs a listener hears.

Specifically, we postulated that reverberant energy would degrade ITD timing information, which would, in turn, blur localization information for all listeners. However, we hypothesized that this blurring might pose special challenges for middle-aged listeners if their ears do not encode fine timing information robustly to begin with. Accordingly, we varied the amount of reverberant energy included in the simulations to vary the task difficulty without changing the task itself. To factor out potential confounds due to other factors known to influence selective attention, we also measured performance on a memory reading span task (e.g., Conway et al. 2001; Akeroyd 2008) and screened out listeners with hearing loss (e.g., Best et al. 2010b; Gatehouse and Noble 2004; Helfer and Freyman 2008). We predicted that performance would decrease as the level of reverberation increased and, after factoring out any effects of hearing thresholds and memory capacity, that performance would decrease with increasing age.

METHODS

Subjects

A total of 33 normal-hearing subjects, ranging in age from 18 to 55 years old, were paid $12.50 per hour for their participation. All gave written informed consent as overseen by the Charles River Campus Institutional Review Board. Roughly dividing the subjects by age, 17 were young adults aged 18–35; 16 were older adults (35–55 years old). All had audiometric thresholds of 20 dB HL or better for octave frequencies from 250 to 8,000 Hz and less than 10 dB of threshold asymmetry between the left and right ears. The young adult subjects were recruited from the Boston University student body. Middle-aged subjects, recruited through word of mouth and posters, were all working professionals from the Boston–Cambridge area. Nearly all held at least bachelor-level, post-secondary degrees.

The hearing thresholds for the younger adults and older adults overlap a great deal (see Fig. 1, where the two light gray regions intersect, forming a dark gray region, over most of the frequency range). The groups' thresholds were well matched, with less than 5 dB HL difference between the mean thresholds at all frequencies (compare thick solid line and thick dashed line in Fig. 1). Although the older adults consistently had greater hearing loss in high frequencies than the younger adults, this difference was not statistically significant (2×7 ANOVA; $p=0.253$).

Working memory task

The Reading Span Test of simultaneous memory storage and processing (Daneman and Carpenter
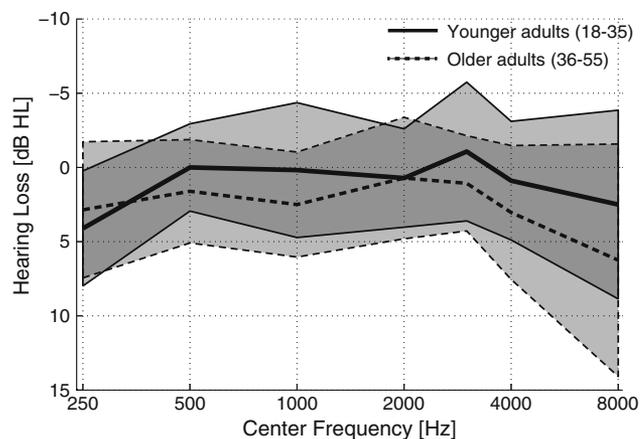
FIG. 1. Younger and older listeners had similar hearing sensitivity. Mean hearing thresholds, ±standard deviation (*shading*), for younger (*solid line*) and older (*dashed line*) listener groups.

1980) was administered to each subject using an established protocol that was previously shown to correlate with performance in an auditory attention task (Lunner 2003). Test sentences were presented in three sequential segments on a computer screen and read aloud by the subject. At the end of the sentence, the subject verbally judged the plausibility of each sentence by responding "yes" if the sentence made sense and "no" if the sentence did not make sense. After all sentences in a block were read, the experimenter requested that the subject verbally report either all of the first words of the sentences in that block or all of the last words of these sentences. There were a total of 12 blocks of sentences presented (three blocks each of three, four, five, and six sentences, corresponding to progressively more challenging blocks). Unit-weighted partial credit scoring (mean of the percent correct within each section) was used to avoid biasing the scores in favor of high performers (Conway et al. 2005).

Spatial selection task

Spatial selective attention was tested by asking listeners to report a sequence of target digits from a virtual location straight ahead in the presence of competing digit streams from two other virtual locations, one 15° to the right and one 15° to the left. The level of reverberant energy was manipulated to modulate task difficulty realistically and without changing other parameters like spatial separation or target to masker ratio. Stimuli in the spatial selection task were processed by binaural room impulse responses (BRIRs) that simulated specular reflections using a rectangular room model implemented in Matlab (Allen and Berkley 1979). The resulting BRIRs thus contained both spatial cues and patterns of reflected

sound energy like those encountered in ordinary spaces.

The modeled room was 7×5 m and 3 m tall. The ears of the simulated listener were located 0.05 m from the center of the room at a height of 1.8 m. The three sources were simulated at a distance of 2.5 m from the receiver; the middle source was directly ahead of the receiver and the two maskers were 15° to the left and 15° to the right of the median plane of the listener. The BRIR was generated by summing, for the direct sound and each subsequent individual reflection, the appropriate head-related impulse response (measured on a KEMAR manikin; see Shinn-Cunningham et al. 2005a), which was time-delayed, filtered, and amplitude scaled as needed, given the wavefront's path from source to the listener's head. Inclusion of this reverberant energy distorted the temporal fine structure in the signals reaching each ear; moreover, because this distortion differs in the two ears, it causes interaural decorrelation.

It is worth pointing out that the use of non-individualized head-related impulse responses can reduce the subjective "realism" in a spatial simulation; however, a lack of realism does not seem to interfere with the ability to use interaural differences for focusing spatial selective attention (e.g., see Shinn-Cunningham et al. 2005b for a discussion of how robust spatial selective attention is even when using severely degraded spatial cues). Moreover, individual differences have little effect on left/right localization (Wenzel et al. 1993) and reverberant energy can mitigate losses in the subjective realism incurred by using non-individualized simulations (Begault et al. 2001). Thus, for the current study of the abilities of listeners to use spatial cues to focus spatial attention in the left/right direction, simulations based on manikin recordings were not expected to yield different results than would have been obtained using individualized simulations.

The room model simulated one of three different levels of wall absorption to achieve three levels of reverberation: anechoic (T60=0 s), intermediate reverberation (T60=0.4 s), and high reverberation (T60=3 s). The absorption coefficients for the three conditions, selected to mimic absorption characteristics of common building materials, are shown in Table 1 for some key frequencies. As noted above, the addition of reverberation altered the temporal fine structure of signals reaching the listeners' ears, introducing interaural decorrelation. This interaural decorrelation interferes with ITDs in the stimuli, resulting in "localization blur" that we expected to reduce the efficacy of spatial selective attention.

The effects of the simulated reverberant energy can be seen in Figure 2, which plots the broadband cross-correlation of the left-ear and right-ear BRIRs

TABLE 1

Absorption coefficients for the three levels of simulated reverberation

Frequency band	125	250	500	1,000	2,000	4,000
Anechoic	1.0	1.0	1.0	1.0	1.0	1.0
Intermediate	0.14	0.10	0.06	0.05	0.09	0.03
High	0.10	0.05	0.06	0.07	0.09	0.08

used in the simulations for sources from left 15°, center, and right 15° (rows) in the anechoic, intermediate reverberation, and high reverberation conditions (columns). In anechoic conditions, the interaural cross-correlation has a single, prominent peak at an ITD corresponding to the appropriate source direction (leftmost column in Fig. 2). However, these peaks are broader and less pronounced in the intermediate reverberation condition (middle column of Fig. 2) and even lower and broader in the high reverberation condition (right column).

This analysis demonstrates why adding reverberant energy may make it difficult to report the target digits, since the target in our task is distinguished from the maskers only by spatial position cues, which are blurred by added reflections. More specifically, because of the small 15° angular separations used, interaural level differences are negligible; the only feature distinguishing target from maskers is the ITD, and this feature is degraded by the interaural decorrelation caused by reflected energy,

Stimuli consisted of single-syllable digits spoken by a single male talker digitally recorded by a sound engineer. The digits included values from one to nine, but excluded the two-syllable digit "seven." Five different tokens of each of the eight digits were used. Digits were convolved with the appropriate BRIRs to simulate the proper, natural spatial cues and reverberant energy content.

On each trial, three simultaneous streams, each containing four spoken digits, were presented. Each of the streams was simulated as coming from one of three directions: straight ahead (the target) or 15° to either side (the two maskers). Digits that were presented simultaneously in the three streams were constrained to differ from each other; the digits presented from a given source at a given time were otherwise selected randomly from among the 40 recorded speech tokens (five tokens of each of eight digits).

Across the three streams, the digits were time aligned so that three digits always started at the same moment, one from each source location. The onset-to-onset separation between consecutive digits from a particular location was 440 ms, which was about the length of the longest raw recorded token; as a result, consecutive digits essentially abutted one another in time in anechoic simulations. In the two echoic conditions, the BRIRs caused significant temporal smearing; thus, the tail of the response to a preceding digit generally overlapped (added) with the beginning of the subsequent digit in these conditions. Specifically, in the intermediate reverberation condition, the average overlap was about 0.4 s; in the highest reverberation condition, the overlap was about 3 s. Due to this temporal overlap from digit to digit, the effects of the reverberation built up throughout the course of each trial.
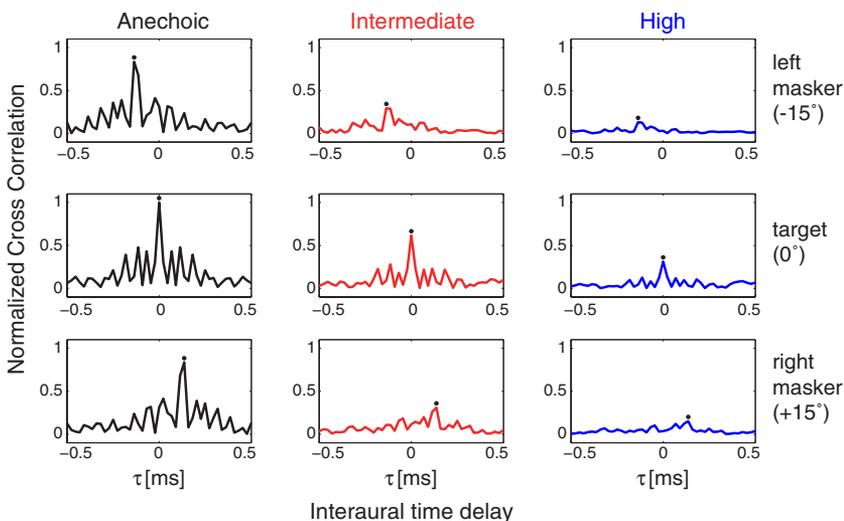
FIG. 2. Reverberant energy causes interaural decorrelation, which is likely to interfere with the ability to selectively attend to the target digits in this spatial selective auditory attention task. Each panel shows the interaural cross-correlation function for a pair of left-ear and right-ear BRIRs used to simulate target (*middle row*) and masker (*top and bottom rows*) streams. Columns correspond to the different levels of simulated reverberation.

After the presentation, listeners were asked to report the sequence of four digits presented from the center stream (the target) using a graphical user interface. At the beginning of the session, the spatial configuration of the three simulated sources was described to the subjects and they were told that they should report the numbers originating from the central source location. They were also made aware that the randomly ordered blocks would differ in their levels of reverberation. Subjects were instructed to guess if they were not sure of the correct response. After this introduction, subjects were given the opportunity to listen to each simulated spatial source in isolation as well as in combination with the other sources (in the anechoic condition). There was no limit on how many times listeners could play these training recordings before testing began; however, no such recordings were available once data collection in the actual experiment began.

Trials were organized into nine blocks: three blocks each of the three reverberant conditions. Blocks were randomly ordered (differently for each subject). Each block consisted of 50 trials, for a total of 150 trials per subject per condition. All nine blocks were performed in one session, which could last up to 3 h. Subjects were encouraged to take breaks as needed. Trials were self-paced: a subsequent trial began after listeners entered their response to the current trial.

RESULTS

Overall, listeners performed best in the anechoic condition, worst in the high-reverberation condition, and in between in the intermediate reverberation condition. Interestingly, when listeners failed to report the target digit, they generally reported one of the simultaneous masker digits, rather than guessing randomly among possible digits. Thus, although completely random responses have a 1/8 chance of being correct, it may be more appropriate to consider the chance of selecting the target digit by chance from among the three digits presented; the chance of being correct by such random *selection* is 1/3.

Of the 33 subjects, 28 subjects reported the target digit more often than expected by random selection chance (33%) in the anechoic (easiest) case. In the easiest, anechoic condition, the remaining five subjects were consistently more likely to report digits from one of the masker streams than to report the target digits. Each of these five listeners consistently chose either the left or right digit stream in the anechoic condition; however, which stream was selected varied from subject to subject. Given these results, we focus on the results of the 28 subjects who were able to perform the task as instructed and

comment briefly on the results of the remaining five subjects in the "Discussion" section.

Figure 3 plots the percent of target digits reported for anechoic, intermediate, and high levels of reverberation (left, center, and right panels of each subfigure, respectively) against age (Fig. 3A), reading span task (Fig. 3B), and pure tone average (Fig. 3C). One of the most striking results is that, in anechoic space (left panels), performance ranged widely across subjects: from greater than 85% down to near 40%. These very large differences in performance in the anechoic condition were not correlated with age ($r=-0.1832$, $p>0.05$; Fig. 3A), reading span score ($r=0.0513$, $p>0.05$; Fig. 3B), or pure tone average ($r=0.078$, $p>0.05$; Fig. 3C).

In the intermediate reverberation condition, performance was lower than in the anechoic condition, ranging from about 30% correct to about 70% correct (compare left and middle panels in Fig. 3A, B, and C). As in the anechoic condition, there were large differences in individual performance; again, these differences were not correlated with age ($r=-0.2035$, $p>0.05$), reading span ($r=-0.0059$, $p>0.05$), or pure tone average ($r=0.093$, $p>0.05$).

In the high reverberation condition, almost every listener performed near or just slightly above chance for selection (near 33%). Nonetheless, a few "star" listeners were able to perform well; one listener reported the target an average of 47% of the time, commensurate with the performance of some of the worst performers in the anechoic condition. Again, neither age nor reading span nor pure tone threshold predicted individual differences in the ability to selectively attend to the target source.

Comparing across the panels in Figure 3, it is clear that on average, performance decreased as the amount of reverberation increased. However, this effect was not only true on average, it was true for every individual listener: each subject performed better in the anechoic condition than in either the intermediate reverberation or the high reverberation conditions.

This is shown in Figure 4, which plots the within-subject difference between anechoic performance and performance in intermediate and high reverberation conditions (left and right bars, respectively). Individual subject data in both the left and right clusters are plotted as symbols against subject age, while the bars show the across-subject means and standard deviations. These data show that performance is always worse when there is more reverberant energy: in every individual case, listeners are better in anechoic conditions than in conditions with reverberation (yielding positive values in Fig. 4). Consistent with the raw results, the difference in performance in the easiest, anechoic condition and in the two
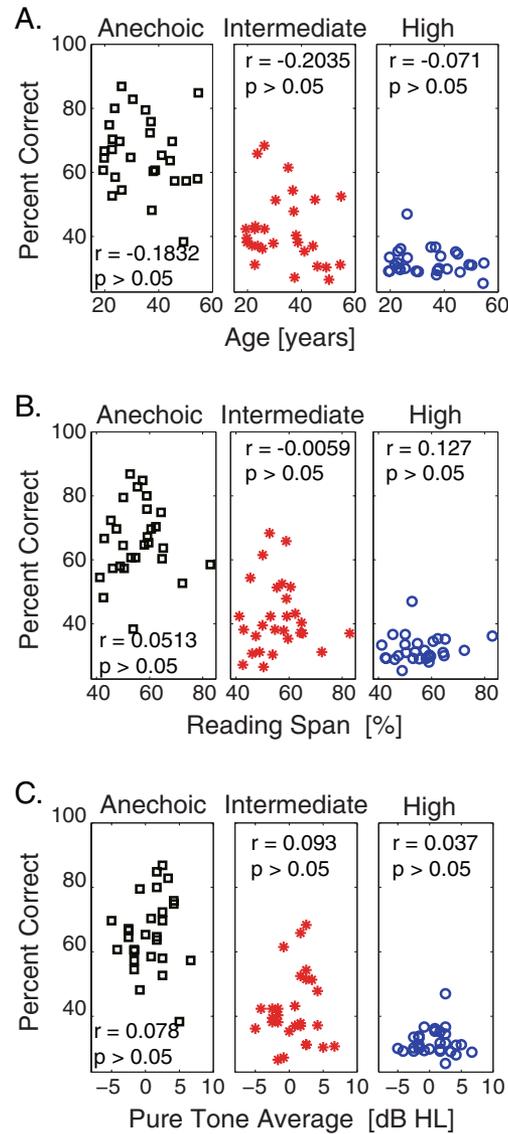
FIG. 3. Performance on the selective attention task was not predicted by age, working memory ability, or average hearing threshold. Scatter plots show percent correct responses in each reverberant condition (left, anechoic; middle, intermediate reverberation; right, high reverberation) for each individual subject. **A** Performance against subject age. **B** Performance against reading span score. **C** Performance against pure tone average.

reverberant conditions was unrelated to age (see Fig. 4; intermediate reverberation: $r=0.024$; $p>0.05$; high reverberation: $r=-0.13$; $p>0.05$) or reading span (not shown; intermediate reverberation: $r=0.095$; $p>0.05$; high reverberation: $r=0.006$; $p>0.05$)

Analysis of responses by digit position

As noted above, response errors were generally not random, but tended to occur because listeners reported one of the masker digits rather than a digit from the central target stream. In addition, there was evidence, at least in anechoic conditions, that listeners became better at reporting the central target from the mixture over the course of a trial, an effect that has been previously observed in similar spatial attention

tasks using multiple competing streams (Best et al. 2008; Best et al. 2010a).

This buildup of spatial selective auditory attention can be seen in the results plotted in Figure 5A, which shows percent correct as a function of the digit position within the four-digit sequence, averaged across the 28 subjects who completed the task as directed.

Specifically, performance was better on digit 4 (average near 80%) than on digit 1 (average near 60%; squares increase from left to right within Fig. 5A). A two-way repeated measures ANOVA showed that digit position had a significant effect on percent correct ($F(3, 324) = 11.53$, $p<0.01$). Reverberation level also had a significant effect on percent correct ($F(2, 324) = 330.37$, $p<0.01$). The significant interaction of digit position and reverberation
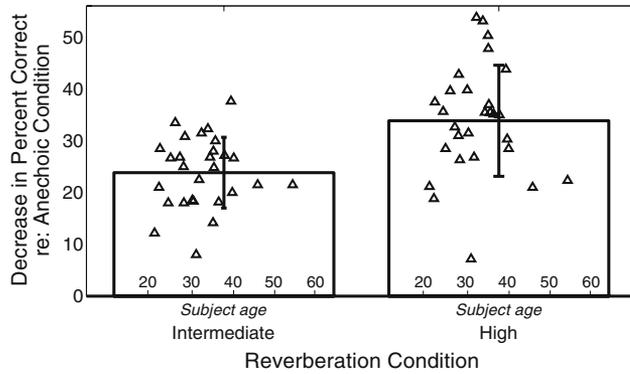
FIG. 4. Each Individual listener performed best in the anechoic condition, an effect that is unrelated to age or other factors. Symbols show the differences between each individual subject's percent correct performance in anechoic and intermediate reverberation conditions (*left*) and in anechoic and high reverberation conditions (*right*), plotted against subject age. Bars show the across-subject means in the difference (*error bars* show the across-subject standard deviation).

$(F(6,324)=63.68, p<0.01)$ supports the observation that build-up is strongest in the anechoic condition, weaker in the intermediate reverberation (see asterisks in Fig. 5A), and completely gone in the high reverberation condition (the circles in Fig. 5A do not vary from left to right within the panel).

Figure 5B, C, and D break down the response errors into reports of left maskers, right maskers, or random guesses (a report of a digit that was not presented in the target stream or either masker stream), respectively, as a function of digit position. Together, these results show that in the high reverberation condition (circles), the average likelihood of reporting the left or right maskers equals the likelihood of reporting the target (all circles in the left three panels fall near 30%). For lower levels of

reverberant energy, the likelihood of reporting the left or right maskers was lower than in the high reverberation condition, an effect that grew stronger over the course of the four-digit long sequence (stars and squares fall below circles in Figure 5B and C, and the distance separating squares from circles increases from left to right within each panel). This decrease in the number of maskers reported from digit to digit supports the idea that attention to the center stream builds up from one digit to the next.

Figure 5D plots the percentage of responses in which listeners guessed, reporting a digit that was not presented in either the target or masker streams. As noted above, these errors happened relatively infrequently, less than 15% of the time in all conditions. Moreover, for digits one and four, these percentages are near zero. Only for digits two and three does the guessing percentage reach levels of 10–15%. These results reflect the traditional serial order phenomena of primacy and recency (e.g., see Jahnke 1965): subjects rarely guessed on the first or last digit in the stream.

The pattern of correct responses (Fig. 5A) in the different conditions reflects a combination of both attentional build-up and serial order (memory) effects. Specifically, attentional build-up (evidenced by a decrease in the likelihood of selecting a masker from digit to digit) causes the probability of a correct response to increase from digit-to-digit. However, primacy and recency effects hurt performance for the middle digits (two and three), resulting in performance that is better on digits one and four. The net result of these two effects is that the probability of being correct increases modestly, at best, from digits one through three (where attentional build-up and primacy/recency effects work in opposition) and then is much higher for digit four (where the two factors work in concert). Thus, even though the memory load is low in our task, we believe that memory effects interact with attentional build-up in determining how the probability of being correct changes from digit to digit.

DISCUSSION

This study sheds insight into the processes that allow the auditory system to focus spatial attention in taxing, reverberant listening conditions. In our spatial selective auditory attention task, listeners rarely guessed the identity of a digit; instead, they almost always reported one of the digits that had been presented (see Fig. 5). This low rate of guessing responses suggests that even the high reverberation did not significantly interfere with the intelligibility of the competing digits. If listeners had been unable to

FIG. 5. Performance improved from one digit to the next; moreover, errors generally corresponded to reports of one of the masker digits, rather than random guesses. Each panel shows the proportion of responses as a function of temporal position within the four-digit stream of each kind of response. **A** Correct responses (*center target*). **B** Left masker responses. **C** Right masker responses. **D** Random guesses (responses of digits not presented in a given time slot).

understand any of the digits, they would have had to guess. Moreover, if reverberant energy had caused significant intelligibility difficulties, they should have been greatest for the final digit and smallest for the initial digit (as reverberation builds up over the course of each trial). Therefore, guessing errors would have been expected to increase with temporal position; instead, guessing responses for the final digit were essentially zero (see Fig. 5D).

The fact that guessing errors are infrequent further suggests that memory was not the primary factor determining how well individual listeners performed. Although memory was not the main limitation on performance, the pattern of guessing errors was consistent with known memory effects. Specifically, in immediate recall tasks, it is common for subjects to perform particularly well when recalling the initial item in a sequence (primacy) and the final item in the sequence (recency; e.g., see Jahnke 1965). These primacy and recency effects account for why we found guessing errors to be greater for the two middle digits than for the initial or final digits. Thus, it seems that in the current four-digit recall task, memory plays a modest, but predictable role.

We believe that the relatively small influence of memory in the current task explains why performance of listeners in our selective attention task was uncorrelated with their Reading Span Score (see Fig. 3B), even though this measure has been seen to correlate with performance in other auditory selective attention tasks (Lunner 2003). In particular, we believe that due to the nature of the speech materials used and the experimental paradigms employed, working memory demands were relatively modest here compared to many other, past studies. Indeed, because our primary interest was in trying to isolate differences in performance that might arise due to early sensory changes in the aging auditory system, rather than cognitive effects, we designed our task to emphasize the ability to encode fine temporal information necessary for computing ITDs and to minimize the influence of memory demands. The lack of memory effects increases the task's specificity in identifying individual differences in performance. The current results strongly suggest that failure of selective spatial attention is the main factor determining performance in our task, rather than intelligibility of the digits or memory constraints; the individual differences in our task reflect differences in the ability to select a desired target stream based on fine ITD cues.

In the anechoic condition, performance tended to improve from digit-to-digit (albeit weakly from digits one through three). Similarly, in the intermediate condition, performance was best on the final digit, even though the reverberant energy increased from digit-to-digit in this condition. Consistent with these effects, the likelihood of reporting a masker digit decreased systematically from digit to digit in both the anechoic and intermediate conditions. This build-up pattern is consistent with a build up of spatial selective focus across time, an effect that has been seen in our laboratory on other, similar tasks (Best et al. 2008; Best et al. 2010a). The presence of this build up lends further support to the idea that selective attention is the primary factor determining performance in the current experiment.

In the current attention task, adding reverberation had a great impact on performance. For each subject who could focus on the correct, target source, adding reverberant energy degraded performance (see Fig. 4). Given that the current task seems to be dominated by demands on selective attention, it thus appears that reverberant energy specifically interferes with spatial selective attention, as expected. Reverberation disrupts temporal fine structure in the signals reaching the listeners' ears; thus, reverberation distorts harmonic structure, interaural correlation, onsets, offsets, and common modulation (e.g., see Nabelek et al. 1989; Darwin and Hukin 2000; Culling et al. 2003), all of which normally aid in segregating competing sources into distinct perceptual objects (Darwin and Carlyon 1995; Shinn-Cunningham 2008). Reverberation is also likely to interfere with higher-order perceptual features like pitch and location, making such features of competing sources perceptually more similar so that listeners cannot select the proper target from the mixture (see Fig. 2 as well as the discussion in Shinn-Cunningham and Best 2008). Indeed, of the minority of subjects who were biased towards reporting one of the side, masker sources rather than the central target, increasing reverberation generally reduced their response bias (i.e., with increasing reverberation, the likelihood of reporting the left masker, right masker, and central target became more equal), suggesting that even listeners who consistently attended to the *wrong* source in anechoic space could not distinguish between target and maskers as well in the presence of reverberant energy (data not shown). Thus, as we originally hypothesized, our results show that reverberant energy disrupts spatial selective auditory attention, consistent with past studies showing that reverberant energy decreases the listening benefit of spatially separating competing sources (Kidd et al. 2005b; Marrone et al. 2008a, b).

Perhaps the most striking finding in the current results is the magnitude of the individual differences in performance in the anechoic and intermediate reverberation conditions. Studies of individual differences done by Watson, et al., have shown large individual differences in general auditory ability, especially in speech processing (Drennan and Watson

2001; Surprenant and Watson 2001; Kidd et al. 2007). While we hypothesized that we would see such differences, we expected them to correlate with age, reflecting degradations in the fidelity with which the sensory system encodes fine temporal details even in pre-senescent (middle-aged) listeners. This expectation was driven by previous studies of middle-aged listeners that reveal deficits in physiological responses evoked by changes in interaural cues (Ross et al. 2007; Ross 2008; Wambacq et al. 2009), as well as psychophysical studies that reveal deficits in performance by middle-aged listeners (Grose et al. 2006; Helfer and Vargo 2009; Grose and Mamo 2010). Despite this, on our task, we found that age did not predict individual differences in performance.

We also found no relationship between performance on our spatial selective attention task and working memory capacity or between performance and thresholds of hearing. As noted above, by design, memory has little effect on performance of our task, which likely explains why Reading Span Score was unrelated to performance on the selective attention task. In addition, because we specifically constrained our subject pool to listeners with clinically normal pure tone hearing thresholds, there was relatively little variation in hearing status across the tested population; given this, it is also reasonable that pure tone averages did not predict selective-attention performance. That said, when divided into groups of younger and older listeners, our older listeners tended to have slightly worse pure tone thresholds at the highest frequencies tested. Thus, as a group, our middle-aged listeners were both older and had slightly worse hearing than our younger listeners. Yet even with both of these factors at play, age was not predictive of performance on the spatial selective auditory attention task (see Fig. 3A).

Although our task was designed to look for early aging effects on the ability to selectively attend, what we found instead were large variations in selective attention ability across our cohort of normal-hearing listeners, whose ages spanned from young adult to middle aged. Normal-hearing thresholds co-occurring with poor performance on spatial listening, listening in noise or multiple-source tasks is often used to diagnose a listener as having auditory processing disorder (APD; see Dawes and Bishop 2009). Thus, tasks with the specificity of the one tested here may be particularly sensitive to identifying individuals with such difficulties.

APD is often assumed to be due to central processing difficulties, not peripheral difficulties. This idea is driven by the fact that APD listeners, by definition, have normal hearing thresholds and because the tasks that are used to diagnose APD often involve higher-order cognitive demands, like requiring a listener to engage spatial selective attention.

However, the fact that our study shows a direct impact of reverberation level on spatial selective auditory attention in normal-hearing listeners calls this logic into question. Specifically, our results show that adding reverberation (a form of peripheral noise) to a scene containing competing speech sources dramatically interferes with the ability to selectively attend to a target from a known direction (a task with strong cognitive demands). This degradation of selective attention was seen even though the reverberation levels we employed left the competing speech signals intelligible. Thus, behavioral deficits on tasks that require attention may reflect differences in how robustly the periphery encodes supra-threshold acoustic inputs (such as differences in the encoding of fine temporal details in the input signals, which are known to be important for selective attention).

A number of recent studies reveal that hearing-impaired listeners often have deficits in temporal fine structure encoding, which can correlate with difficulties in understanding speech in noise (see Lorenzi et al. 2006; Strelcyk and Dau 2009). However, our results hint that many "normal-hearing" listeners may have deficits in encoding of temporal fine structure in supra-threshold signals, even though their thresholds of hearing are not affected. Recent animal studies reveal that noise exposure that does not permanently alter pure tone thresholds can nonetheless be accompanied by a loss of spiral ganglion cells (Kujawa and Liberman 2009). Given that convergence of multiple auditory nerve fibers is thought to lead to the exquisite precision of temporal fine structure encoding in the cochlear nucleus (e.g., see Joris and Smith 2008), loss of spiral ganglion is likely to result in a reduction in the encoding of temporal fine structure in supra-threshold signals. Moreover, listeners with normal pure tone thresholds can show large differences in the reliability with which these brainstem encodes the harmonic structure of speech sounds, which correlates with their ability to understand speech in noise (e.g., Anderson et al. 2010; Skoe and Kraus 2010). Finally, studies of temporal jitter and spectral smearing suggest that processing of temporal fine structure in low-frequency signals is perceptually important (MacDonald et al. 2010). Taken together, these studies suggest that a cohort of listeners with similar, "normal" audiometric thresholds may nonetheless have large differences in their supra-threshold peripheral auditory function. Such variations may help account for the differences we see in normal-hearing listeners' ability to selectively attend to a target source among similar, competing speech streams.

Future work to explore the relationship between peripheral encoding and performance on "central" tasks like the spatial attention task employed here is

critical to address this possibility. For instance, comparing individuals' ability to detect fine temporal information and their performance on our spatial selective attention task may reveal a systematic correlation (e.g., see Strelcyk and Dau 2009). Physiological measures of brainstem encoding of fine temporal structure could also correlate with spatial selective attention performance (e.g., Anderson et al. 2010; Skoe and Kraus 2010). Such relationships could yield insight into the sources of individual differences in selective attention; we are currently conducting follow-up experiments with some our normal-hearing listeners to explore this possibility.

By design, the current task emphasizes the importance of encoding fine temporal information in focusing spatial selective attention; as discussed above, we crafted our task to be sensitive to differences in supra-threshold sensory coding in the periphery of the auditory system. We find that there is a broad range of ability on our task. It is likely that, if they were seen in the clinic, some of our worst listeners would be diagnosed as having APD or some similar syndrome. There is undoubtedly a range of factors that contribute to perceptual difficulties leading to diagnoses of APD, including deficits in cognitive function, executive control, and sensory coding (including encoding of temporal fine structure). However, the very specificity of the current spatial selective attention task may prove useful in the clinic as a way to measure the contribution of one specific factor to the ability to communicate in everyday settings. By isolating specific mechanisms that can lead to different clinical diagnoses (such as APD), it may be possible to develop more specific and effective treatments and devices to aid listeners with such problems.

ACKNOWLEDGMENTS

REFERENCES

Akeroyd MA (2008) Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. Int J Audiol 47(Suppl 2):S53–S71

Allen JB, Berkley DA (1979) Image method for efficiently simulating small-room acoustics. J Acoust Soc Am 65(4):943–950

Anderson S, Skoe E, Chandrasekaran B, Kraus N (2010) Neural timing is linked to speech perception in noise. J Neurosci 30(14):4922–4926

Arlinger S, Lunner T, Lyxell B, Pichora-Fuller MK (2009) The emergence of cognitive hearing science. Scand J Psychol 50(5):371–384

Begault DR, Wenzel EM, Lee AS, Anderson MR (2001) Direct comparison of the impact of head-tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. J Audio Eng Soc 49(10):904–916

Best V, Ozmeral EJ, Kopco N, Shinn-Cunningham BG (2008) Object continuity enhances selective auditory attention. Proc Natl Acad Sci 105(35):13174–13178

Best V, Shinn-Cunningham BG, Ozmeral EJ, Kopco N (2010a) Exploring the benefit of auditory spatial continuity. J Acoust Soc Am 127(6):EL258–EL264

Best V, Gallun FJ, Mason CR, Kidd G Jr, Shinn-Cunningham BG (2010b) The impact of noise and hearing loss on the processing of simultaneous sentences. Ear Hear 31(2):213–220

Bronkhorst AW (2000) The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions. Acustica 86:117–128

Cherry EC (1953) Some experiments on the recognition of speech, with one and with two ears. J Acoust Soc Am 25:975–979

Conway AR, Cowan N, Bunting MF (2001) The cocktail party phenomenon revisited: the importance of working memory capacity. Psychon Bull Rev 8(2):331–335

Conway AR, Kane MJ, Bunting MF, Hambrick DZ, Wilhelm O, Engle RW (2005) Working memory span tasks: a methodological review and user's guide. Psychon Bull Rev 12(5):769–786

Culling JF, Hodder KI, Toh CY (2003) Effects of reverberation on perceptual segregation of competing voices. J Acoust Soc Am 114(5):2871–2876

Daneman M, Carpenter P (1980) Individual differences in working memory and reading. J Verbal Learn Verbal Behav 19(4):450–466

Darwin CJ, Carlyon RP (1995) Auditory grouping. Hearing. B. C. J. Moore. San Diego, CA, Academic Press, pp 387–424

Darwin CJ, Hukin RW (2000) Effects of reverberation on spatial, prosodic, and vocal-tract size cues to selective attention. J Acoust Soc Am 108(1):335–342

Darwin CJ, Brungart DS, Simpson BD (2003) Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. J Acoust Soc Am 114(5):2913–2922

Dawes P, Bishop D (2009) Auditory processing disorder in relation to developmental disorders of language, communication and attention: a review and critique. Int J Lang Commun Disord 44(4):440–465

Drennan WR, Watson CS (2001) Sources of variation in profile analysis. I. Individual differences and extended training. J Acoust Soc Am 110(5 Pt 1):2491–2497

Francis AL (2010) Improved segregation of simultaneous talkers differentially affects perceptual and cognitive capacity demands for recognizing speech in competing speech. Atten Percept Psychophys 72(2):501–516

Freyman RL, Helfer KS, McCall DD, Clifton RK (1999) The role of perceived spatial separation in the unmasking of speech. J Acoust Soc Am 106(6):3578–3588

Gatehouse S, Noble W (2004) The Speech, Spatial and Qualities of Hearing Scale (SSQ). Int J Audiol 43(2):85–99

Grose JH, Mamo SK (2010) Processing of temporal fine structure as a function of age. Ear Hear 31:755–760

Grose JH, Hall JW 3rd, Buss E (2006) Temporal processing deficits in the pre-senescent auditory system. J Acoust Soc Am 119(4):2305–2315

Grose JH, Mamo SK, Hall JW 3rd (2009) Age effects in temporal envelope processing: speech unmasking and auditory steady state responses. Ear Hear 30(5):568–575

Helfer KS, Freyman RL (2008) Aging and speech-on-speech masking. Ear Hear 29(1):87–98

Helfer KS, Vargo M (2009) Speech recognition and temporal processing in middle-aged women. J Am Acad Audiol 20(4):264–271

Hill KT, Miller LM (2010) Auditory attentional control and selection during cocktail party listening. Cereb Cortex 20(3):583–590

Ihlefeld A, Shinn-Cunningham B (2008) Disentangling the effects of spatial cues on selection and formation of auditory objects. J Acoust Soc Am 124(4):2224–2235

Jahnke JC (1965) Primacy and recency effects in serial-position curves of immediate recall. J Exp Psychol 70:130–132

Joris PX, Smith PH (2008) The volley theory and the spherical cell puzzle. Neuroscience 154(1):65–76

Kidd G Jr, Arbogast TL, Mason CR, Gallun FJ (2005a) The advantage of knowing where to listen. J Acoust Soc Am 118(6):3804–3815

Kidd G Jr, Mason CR, Brughera A, Hartmann WM (2005b) The role of reverberation in release from masking due to spatial separation of sources for speech identification. Acustica United with Acta Acustica 91:526–536

Kidd GR, Watson CS, Gygi B (2007) Individual differences in auditory abilities. J Acoust Soc Am 122(1):418–435

Kujawa SG, Liberman MC (2009) Adding insult to injury: cochlear nerve degeneration after "temporary" noise-induced hearing loss. J Neurosci 29(45):14077–14085

Lavandier M, Culling JF (2007) Speech segregation in rooms: effects of reverberation on both target and interferer. J Acoust Soc Am 122(3):1713

Lavandier M, Culling JF (2008) Speech segregation in rooms: monaural, binaural, and interacting effects of reverberation on target and interferer. J Acoust Soc Am 123(4):2237–2248

Lorenzi C, Gilbert G, Carn H, Garnier S, Moore BC (2006) Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. Proc Natl Acad Sci USA 103(49):18866–18869

Lunner T (2003) Cognitive function in relation to hearing aid use. Int J Audiol 42(Suppl 1):S49–S58

MacDonald EN, Pichora-Fuller MA, Schneider BA (2010) Effects on speech intelligibility of temporal jittering and spectral smearing of the high-frequency components of speech. Hearing Research 261:63–66

Marrone N, Mason CR, Kidd G (2008a) Tuning in the spatial dimension: evidence from a masked speech identification task. J Acoust Soc Am 124(2):1146–1158

Marrone N, Mason CR, Kidd G Jr (2008b) The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms. J Acoust Soc Am 124(5):3064–3075

Nabelek AK, Letowski TR, Tucker FM (1989) Reverberant overlap- and self-masking in consonant identification. J Acoust Soc Am 86:1259–1265

Pichora-Fuller MK, Singh G (2006) Effects of age on auditory and cognitive processing: implications for hearing aid fitting and audiologic rehabilitation. Trends Amplif 10(1):29–59

Ross B (2008) A novel type of auditory responses: temporal dynamics of 40-Hz steady-state responses induced by changes in sound localization. J Neurophysiol 100(3):1265–1277

Ross B, Fujioka T, Tremblay KL, Picton TW (2007) Aging in binaural hearing begins in mid-life: evidence from cortical auditory-evoked responses to changes in interaural phase. J Neurosci 27(42):11172–11178

Sheldon S, Pichora-Fuller MK, Schneider BA (2008a) Effect of age, presentation method, and learning on identification of noise-vocoded words. J Acoust Soc Am 123(1):476–488

Sheldon S, Pichora-Fuller MK, Schneider BA (2008b) Priming and sentence context support listening to noise-vocoded speech by younger and older adults. J Acoust Soc Am 123(1):489–499

Shinn-Cunningham BG (2008) Object-based auditory and visual attention. Trends Cogn Sci 12(5):182–186

Shinn-Cunningham BG, Best V (2008) Selective attention in normal and impaired hearing. Trends Amplif 12(4):283–299

Shinn-Cunningham BG, Ihlefeld A, Satyavarta LE (2005a) Bottom-up and top-down influences on spatial unmasking. Acta Acustica United with Acustica 91(6):967–979

Shinn-Cunningham BG, Kopco N, Martin TJ (2005b) Localizing nearby sound sources in a classroom: binaural room impulse responses. J Acoust Soc Am 117(5):3100–3115

Singh G, Pichora-Fuller MK, Schneider BA (2008) The effect of age on auditory spatial attention in conditions of real and simulated spatial separation. J Acoust Soc Am 124(2):1294–1305

Skoe E, Kraus N (2010) Auditory brain stem response to complex sounds: a tutorial. Ear Hear 31(3):302–324

Snell KB, Mapes FM, Hickman ED, Frisina DR (2002) Word recognition in competing babble and the effects of age, temporal processing, and absolute sensitivity. J Acoust Soc Am 112(2):720–727

Strelcyk O, Dau T (2009) Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing. J Acoust Soc Am 125(5):3328–3345

Surprenant AM, Watson CS (2001) Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners. J Acoust Soc Am 110(4):2085–2095

Tun PA, O'Kane G, Wingfield A (2002) Distraction by competing speech in young and older adult listeners. Psychol Aging 17(3):453–467

Wambacq IJ, Koehnke J, Besing J, Romei LL, Depierro A, Cooper D (2009) Processing interaural cues in sound segregation by young and middle-aged brains. J Am Acad Audiol 20(7):453–458

Wenzel EM, Arruda M, Kistler DJ, Wightman FL (1993) Localization using nonindividualized head-related transfer functions. J Acoust Soc Am 94(1):111–123

Zurek PM, Freyman RL, Balakrishnan U (2004) Auditory target detection in reverberation. J Acoust Soc Am 115(4):1609–1620