Coherence of the irrelevant-sound effect: Individual profiles of short-term memory and susceptibility to task-irrelevant materials

EMILY M. ELLIOTT
Louisiana State University, Baton Rouge, Louisiana

and

NELSON COWAN
University of Missouri, Columbia, Missouri

We examined individual and developmental differences in the disruptive effects of irrelevant sounds on serial recall of printed lists. In Experiment 1, we examined adults ($N = 205$) receiving eight-item lists to be recalled. Although their susceptibility to disruption of recall by irrelevant sounds was only slightly related to memory span, regression analyses documented highly reliable individual differences in this susceptibility across speech and tone distractors, even with variance from span level removed. In Experiment 2, we examined adults ($n = 64$) and 8-year-old children ($n = 63$) receiving lists of a length equal to a predetermined span and one item shorter (span $-$ 1). We again found significant relationships between measures of span and susceptibility to irrelevant sounds, although in only two of the measures. We conclude that some of the cognitive processes helpful in performing a span task may not be beneficial in the presence of irrelevant sounds.

The disruptive effects of irrelevant sounds on immediate serial recall are of considerable interest. This disruption is typically referred to as the *irrelevant-speech effect* (Colle & Welsh, 1976; Salamé & Baddeley, 1982) but also the *irrelevant-sound effect* (ISE; Jones & Macken, 1993), given that it can occur with nonspeech sounds. Investigators have noted theoretical implications of this task for understanding memory, attention, and perception (for a recent review, see Neath, 2000, and commentaries by Baddeley, 2000, and Jones & Tremblay, 2000). Fundamentally, the ISE reflects an important limitation in humans' ability to determine which stimuli will enter into the mnemonic processing needed for short-term recall. To the extent that the ISE affects processes within short-term recall that are also used more widely in cognitive tasks, it is of broad significance. Yet important questions remain.

A key issue we will address is whether an individual's degree of susceptibility to the ISE is materials specific or general. Jones and Macken (1993) found that either speech or tones could disrupt serial recall, as long as the auditory stimuli kept changing perceptibly (e.g., the use of multiple syllables or tone pitches, as opposed to a single repeating stimulus). According to their theoretical account (developed by Jones, 1993), the disruption occurs for the same reason for irrelevant speech sounds or tones: because the sounds corrupt the formation of a short-term episodic record.

We do not know of a theory asserting that speech and tone interference do not share a common process. However, a theory of Neath (2000) asserts that speech interference occurs primarily because features from speech are adopted into the phonological memory record of the printed items to be recalled. To account for irrelevant-tone effects, an additional assumption was that changes in irrelevant sounds can cause distraction (cf. Cowan, 1995).

In light of Neath's (2000) approach, speech-based and tone-based ISEs theoretically could occur for largely different reasons (speech effects, because of feature adoption and tone effects, because of distraction). Different sets of individuals might be more susceptible to speech interference than to tone interference, and the correlation between them would be low. In contrast, a strong correlation between speech-based and tone-based ISEs, indicating that some individuals are more susceptible to both kinds of interference, would help to constrain Neath's model by suggesting that the same mechanism should be

used to account for much of the variance in both types of ISE (as posited also by Jones, 1993). We address this question with a large sample of adults and children, using regression analyses.

A second question we will address (in addition to the materials-specific versus general basis of disruption by irrelevant sounds) concerns the role of attentional distraction. If distraction is a component of the ISE (as suggested by Cowan, 1995; see also Buchner, Rothermund, Wentura, & Mehl, 2004), one would expect that individuals with better attention-related skills would be better able to avoid distraction. One way that attention-related skills have been examined in the previous literature is through the measurement of working-memory (WM) capacity, in which individuals engage in tasks that require some elements of additional processing while attempting to retain information for later recall (Daneman & Carpenter, 1980). Given that individuals with better WM abilities appear to be better at using attention to resist interference (e.g., Kane & Engle, 2003), one could similarly expect individuals with better serial recall or working memory to demonstrate smaller ISEs.

In contrast to this second expectation, though, two studies have shown that correlations between irrelevant-speech effects and serial recall are nonsignificant despite large individual differences in susceptibility to irrelevant speech (Ellermeier & Zimmer, 1997; Neath, Farley, & Surprenant, 2003). A third study (Beaman, 2004) similarly indicates no significant correlation between irrelevant speech and the level of performance in an often-used measure of WM capacity known as *operation span*, which taxes attention by requiring participants to carry out an arithmetic operation before each item to be retained for later recall (Turner & Engle, 1989).

The absence of these correlations in adults stands in contrast, however, to what has been found in a developmental study with children and adults (Elliott, 2002). Elliott obtained the well-established increase in memory span with age and a significant decrease in the magnitude of the ISE, which was especially striking for speech interference (i.e., the ISE was particularly large in children compared with adults). The results of this study were interpreted with respect to the role of attention in the ISE. As children develop, according to the interpretation, they gain better control of their attentional abilities, leading to a reduction in the disruption shown by irrelevant sounds.

According to certain assumptions, one would also expect negative correlations between span and the magnitude of the ISE within an age group. The key such assumption is that the factors influencing span and the ISE are the same or are correlated with one another within an age group. The link to the developmental evidence depends on the additional assumptions that developmental changes in span (and in the magnitude of the ISE) depend on changes in the same factors that are important in distinguishing between individuals within an age group. Given the developmental evidence, accepting these assumptions leads to a theory in which individuals with higher spans should be less susceptible to the ISE than those with lower spans. One can see, though, several ways in which the assumptions could be wrong. For example, development includes not only an improvement in attention (Lane & Pearson, 1982; Zukier & Hagen, 1978), but also in covert verbal rehearsal (Cowan & Kail, 1996; Flavell, Beach, & Chinsky, 1966; Ornstein & Naus, 1978). It is always possible that span and diminution of the ISE depend on attention and rehearsal in different ways or that these two factors play different roles in developmental versus individual-difference data. These assumptions are worth exploring, and the absence of significant correlations in adult studies therefore warrants further investigation.

We reexamined these correlations with modifications of previously used methods. We examined correlations between serial recall performance and the effect of irrelevant tones, whereas previous studies examined only correlations using speech interference. If a speech-specific process contributes a great deal to the effect of the ISE, as some have theorized (Neath, 2000; Salamé & Baddeley, 1982), and if this speech-specific process is unrelated to attention, it might obscure a smaller, attention-related effect that might correlate with span. With interfering tones, any obscuring effect of a speech-specific process would be eliminated, allowing a clearer examination of a possible correlation between an attention-related process and level of performance in serial recall.

Also, in correlational studies using serial recall, the measures of performance were somewhat unconventional. In the study of Ellermeier and Zimmer (1997), no correlation between the ISE and memory capacity was obtained ($N = 72, r = .01$). However, the measure of capacity was nonstandard. Given that recall in a quiet control condition was used to calculate the magnitude of the ISE, a separate, pink-noise condition was used as the measure of memory capacity. In the study of Neath et al. (2003), the quiet control condition was used for the performance measure and a proportional measure of the ISE was used. Specifically, the speech–quiet difference score was divided by the performance level in quiet and multiplied by 100 to produce a percentage effect. With $N = 100$, the percentage effect of irrelevant speech was found not to correlate with the level of performance in quiet. However, in neither of these studies was the measure of performance a separate memory-span task. In our Experiment 1, it was.

EXPERIMENT 1
Memory Span and Irrelevant Sounds

This large-scale experiment comprises six subexperiments (1A–1F) conducted originally for a different purpose. We had been trying to understand a finding that the disruptive effect of tones presented along with the visual

items to be recalled occurred only sporadically in procedures that also included a long retention interval between the stimuli and the recall cue. In the process of investigating that finding—which we still do not understand—we ran control experiments in which no retention interval was included. These showed ISEs with both speech and tones and were combined here to allow a large-scale correlational investigation. The questions under investigation were (1) whether the speech- and tone-based ISEs were related, and (2) whether there were at least modest correlations between ISEs and memory span.

Method

Participants

A total of 211 participants were originally assessed, but 2 males and 2 females reported hearing loss. One female reported taking pain medicine, and a computer problem excluded another female. In the final sample, $N = 205$ (157 female, 48 male). The participants were enrolled in psychology classes at the University of Missouri and Louisiana State University. All reported normal or corrected-to-normal vision and hearing and were native speakers of English.

Apparatus, Stimuli, and Procedure

Overview. Six subexperiments (1A–1F) all featured a fixed list length of eight items with immediate recall. The only differences among the subexperiments were in how the tones were selected and how many tones were used on each trial (see below). None of the analyses indicated any differences in the disruption caused by the variations of the tone conditions. This is consistent with previous research by Tremblay and Jones (1998) showing that the "token set size" does not increase the amount of disruption when the number of tokens is larger than two.

The participants were seated in a quiet room or booth for an immediate visual-span assessment, a delayed visual-span assessment, and then the irrelevant-sound task. In all three tasks, lists of digits from the set 1–9 (randomly selected without replacement for each list) were presented on a computer screen for immediate recall. The digits in a list were presented one at a time in the center of the screen, in 30-point font, at the rate of one per second, with no interstimulus interval. The participants' responses were typed using the computer keypad as described below.

Immediate span task. The participants first carried out three practice trials with three-digit lists, followed by test lists that began with three digits per list. Four lists were presented at each list length, with length increasing until the point at which a participant failed to answer at least one list out of the four correctly. The range of list lengths was from three to nine items.

A fixation cross appeared in the center of the screen for 750 msec prior to the first digit appearing. After the set of digits was presented for a particular list, a row of white line segments appeared on the screen, cuing the participants to type in the response. The number of lines matched the number of items presented in a given list. When the response was completed, the computer queried whether a change was needed, because the participants were not allowed to backspace during digit entry. If they responded "yes" by pressing the "y" key, they were given a chance to change their responses. The participants responded "no" by pressing the "n" key. After the prompt to indicate *yes* or *no* (and possible digit entry), the computer continued on to the next list. The presentation of the next list was initiated by the participant pressing the space bar, and the next trial began immediately. Instructions for this task were to remember the numbers in the order in which they were presented, and to guess if unsure. The participants were told not to speak the numbers aloud or to whisper to themselves, but to "think about the numbers inside your head."

Delayed visual-span task. This task was identical to the immediate span task with one exception. After seeing the digits, the participants waited during a retention interval before they were cued to respond. The duration of the retention interval was equal to that for the visual list presentation. For example, if participants were shown a six-item list for recall, they then saw a black screen for 6 sec before the recall cue (white lines) appeared. The instructions were to recall the numbers in order, as in the previous task. In addition, they were told that extra time would be given before they were asked to provide their answer and they were encouraged to think actively about the numbers during this time.

Serial recall with irrelevant sounds. Each trial included an eight-digit list, visually presented. In Subexperiments 1A, 1B, and 1C, there were three auditory conditions (words, tones, and silence). The participants first completed 3 trials in a practice block, and then nine test blocks of 3 trials each, for a total of 30 trials. In Subexperiments 1D, 1E, and 1F, there were four auditory conditions (words, four tones, eight tones, and silence), and there were 4 trials in a practice block followed by nine test blocks of 4 trials each, for a total of 40 trials. In all experiments, the auditory conditions were randomly ordered within each block of trials.

Further details about the subexperiments are as follows. In Subexperiment 1A ($n = 30$), University of Missouri students heard eight tones out of a possible selection of nine on each irrelevant-tone trial, selected randomly without replacement. In Subexperiments 1B and 1C (total $n = 73$), Louisiana State University students heard four of the tones, randomized and repeated twice during the visual presentation of eight digits to be recalled. Subexperiments 1D (at Missouri) and 1E and 1F (at LSU; total $n = 101$) included both the four-tone and the eight-tone conditions. In each of these experiments, the conditions with four tones and with eight tones were not significantly different in the amount of disruption to serial recall performance. For correlational purposes, therefore, the average across the two tone conditions was used in these subexperiments.

The irrelevant spoken words were selected randomly without replacement from the following: *red*, *blue*, *green*, *yellow*, *white*, *tall*, *big*, *short*, and *long*. These words were digitally recorded from a male speaker and ranged in duration from 210 to 500 msec. The sine-wave tones were 500 msec in duration and were selected randomly from the following set of frequencies: 87, 174, 266, 348, 529, 696, 788, 880, and 972 Hz. These were the same words and tones used by Elliott (2002), affording a comparison across studies.

All of the sounds in a trial were presented with a 1-sec onset-to-onset period, with onsets synchronized with the visual stimuli. The sounds were presented over audiological headphones, and sound levels, measured with a Quest sound-level meter and earphone coupler, all fell within the range of 62–68 dB(A).

The instructions in this task were to concentrate on remembering the digits in order and to ignore any sounds that were heard during any portion of the task. The participants were asked not to say anything aloud during the task. Halfway through the task, they were offered a break.

Results and Discussion

Means

The means for every subexperiment and for the full sample are presented in Table 1. Two measures of span were calculated to assess the number of lists answered correctly at each list length. The cumulative span measure was created by finding the highest list length at which a participant answered all four lists correctly, and then adding .25 for each additional list above that list length (see Elliott, 2002). The second measure, referred to as *maximum span*, was calculated as the highest list length at which at least one list was answered correctly.

Table 1
Means and Standard Errors in the Six Subexperiments Within Experiment 1, Separately and Together

	Subexperiment													
	1A		1B		1C		1D		1E		1F		All (1A–1F)	
Measure	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE
Span task														
DelCum	7.4	0.20	6.5	0.13	6.6	0.16	6.5	0.19	6.7	0.12	6.7	0.14	6.7	0.07
DelMax	8.3	0.19	7.6	0.18	7.6	0.16	7.5	0.22	7.7	0.15	7.8	0.15	7.7	0.07
ImmCum	7.1	0.19	6.2	0.17	6.3	0.13	6.3	0.18	6.7	0.16	6.5	0.16	6.5	0.07
ImmMax	7.8	0.20	7.3	0.20	7.4	0.18	7.3	0.20	7.9	0.19	7.4	0.17	7.5	0.08
List-recall task														
Silence	0.80	0.03[b]	0.64	0.03[c]	0.68	0.02[c]	0.64	0.04[b]	0.73	0.03[c]	0.73	0.02[c]	0.71	0.01[c]
Speech	0.73	0.04[a]	0.52	0.03[a]	0.57	0.03[a]	0.56	0.04[a]	0.59	0.03[a]	0.58	0.03[a]	0.59	0.01[a]
Four tones			0.59	0.03[b]	0.64	0.03[b]	0.63	0.03[b]	0.66	0.03[b]	0.66	0.02[b]		
Eight tones	0.78	0.03[b]					0.60	0.03[b]	0.67	0.03[b]	0.65	0.02[b]		
All tones	0.78	0.03	0.59	0.03	0.64	0.03	0.62	0.03	0.66	0.03	0.65	0.02	0.66	0.01[b]

Note—Span task measures: DelCum, delayed test, cumulative score; DelMax, delayed test, maximum score; ImmCum, immediate test, cumulative score; ImmMax, immediate test, maximum score. Within each of the six subexperiments, list-recall condition scores marked with different superscript letters (a–c) are significantly different in Newman–Keuls post hoc tests. 1A, $n = 31$; 1B, $n = 30$; 1C, $n = 43$; 1D, $n = 31$; 1E, $n = 30$; 1F, $n = 40$; total $N = 205$.

Correlations

We followed the previous literature on individual differences in the ISE, creating difference scores to compare the magnitude of the irrelevant-speech effect and the irrelevant-tone effect. In each case, the sound condition performance level was subtracted from the level of performance in the silent control condition. These derived variables were approximately normally distributed.

The two indices were entered into a correlational analysis with the four measures of span, the silent control condition, the speech condition, and the tone condition (shown in Table 2). The generally high correlations between different span scores demonstrates that they were reliable. The generally low correlations between the ISE effects and the immediate spans is in keeping with past results (Ellermeier & Zimmer, 1997; Neath et al., 2003).

To evaluate the reliability of all of the variables, separate reliability analyses were conducted following the methodology of Ellermeier and Zimmer (1997). On the basis of the raw measures, all of which had reliabilities of .81 and above, proportion correct values from adjacent speech and silence trials were used to create speech-effect difference scores for each participant, resulting in nine difference scores per participant. These nine scores were used to calculate Cronbach's alpha with a resulting reliability estimate ($\alpha = .33$). In comparison, Ellermeier and Zimmer had a higher number of trials per participant than the present study, which resulted in 20 scores for each of their 72 participants and a higher reliability estimate ($\alpha = .55$).

Each participant also contributed nine trials in the tone condition (for the experiments containing more than one type of tone condition, the average of the tone trials was used). The same method was used to calculate the reliability of the tone-effect difference score ($\alpha = .20$). Given the low reliability of the difference scores, below we examine our hypotheses with regressions based on raw scores, which were much more reliable, as is shown in Table 2.

Unlike in the past studies, there were significant, albeit modest, correlations—in particular, between the maximum measure of delayed span and both the speech effect and the tone effect ($r = .17$ and $.16$, respectively, with $N = 205$; with this number of participants, $r = .14$

Table 2
Correlations Between Span Measures and ISE Measures in Experiment 1

	1	2	3	4	5	6	7	8	9
DelCum (1)	.84	.97	.81	.66	.26	.10	.89	.76	.79
DelMax (2)	.81*	.82	.72	.62	.40	.33	.83	.61	.69
ImmCum (3)	.68*	.60*	.84	1.00	.12	.00	.81	.72	.75
ImmMax (4)	.55*	.51*	.85*	.83	.30	.00	.70	.63	.59
ToneEffect (5)	.10	.16*	.05	.12	.20	1.00	.81	−.27	−.76
SpeechEffect (6)	.05	.17*	.00	.00	.59*	.33	.47	−.89	−.25
Silence (7)	.72*	.67*	.66*	.57*	.32*	.24*	.81	.93	.98
Speech (8)	.63*	.50*	.60*	.52*	−.11	−.46*	.75*	.82	.99
Tones (9)	.66*	.57*	.63*	.49*	−.31*	−.13	.80*	.82*	.84

Note—DelCum, delayed test, cumulative score; DelMax, delayed test, maximum score; ImmCum, immediate test, cumulative score; ImmMax, immediate test, maximum score. Scores on the diagonal (in italic) represent Cronbach's alpha reliability coefficients. Scores above the diagonal represent the correlations after correction for attenuation. These were not evaluated for statistical significance.   *$p < .05$.

would be statistically significant). The direction of correlation was unanticipated, in which individuals with a higher span had larger ISEs. However, these correlations became nonsignificant when the 4 individuals with the lowest delayed maximal spans (five items) were removed from the sample. One possible interpretation of the effect is that the levels of serial-recall performance in these 4 individuals were so low in the silent condition ($M = .40$, $SD = .09$, as opposed to a much higher mean for the entire sample, $M = .71$, $SD = .17$) that they reflected only a basic form of memory that is resistant to disruption (e.g., the focus of attention; see Cowan, 2001).

We analyzed our ability to detect correlations in this sample, using $N = 201$ after the 4 individuals with the lowest delayed maximal spans were removed. Without those participants, the correlation between the maximum measure of delayed span and the speech effect was $r = .10$, and the resulting power was low (power $= .29$). To obtain a more acceptable level of power, such as the standard .80, a minimum correlation ($r = .20$) would be required.

Regressions

The high correlation between the speech and tone effects ($r = .59$; see Table 2) is potentially of central interest but cannot be cleanly interpreted, given that both effects were calculated using the same silent control condition. To overcome this problem, we entered the ISE variables and silent control condition into stepwise regression analyses, as seen in Table 3.

In the first such analysis, the dependent variable was the tone effect. As the table shows, a significant amount of variance was shared by the tone effect and the silent control condition, which was entered first into the regression equation. What is important in the analysis is that the speech effect, entered after the silent condition, still accounted for very substantial and significant additional variance in the tone effect ($\Delta R^2 = .28$). Similarly, when the speech effect was the dependent variable and the silent control condition was removed first, the tone effect still accounted for substantial and significant additional variance ($\Delta R^2 = .29$).

These regressions indicate that susceptibilities to the speech and tone effects are closely related, for reasons having nothing to do with the use of a common control condition to calculate the effects. Almost identical results were obtained when we omitted the 4 outlier participants with a delayed maximal span of only five items (see above).

Figure 1 further illustrates, quite explicitly, that the susceptibility to irrelevant sounds is unrelated to memory span. It is a scatterplot of the irrelevant-tone effect as a function of the irrelevant-speech effect, with different symbols to indicate the immediate span (maximal measure) for each individual. There is clearly a strong relation between irrelevant-speech and irrelevant-tone effects, and for each span level there is a range of susceptibility levels across individuals.

Inasmuch as the reliability estimates for the speech and tone effects were much lower than the estimates for the raw scores, additional regression analyses were conducted using the raw scores from the speech and tone conditions in place of the difference scores. The results are shown in Table 4. The pattern is very similar to the analyses using the difference scores. Variability in the irrelevant-tone condition was predicted first by the silent control condition; and the irrelevant-speech condition, entered after the silent condition, accounted for significant additional variance. Similarly, the irrelevant-tone condition predicted significant variance in the irrelevant-speech condition even after the relation to the silent condition was taken out.

We also carried out regressions on each type of span measure using serial recall performance in the presence of silence, speech, or tones as the predictors (see Table 5). There were some significant regression effects of the irrelevant-sound conditions after the silent-condition variation was taken out, but they accounted for only very small amounts of variance in the span measures.

Table 3
Regression Analyses in Experiments 1 and 2 Using Difference (Effect) Scores

Age Group	Dependent Measure	Regression Step 1	Regression Step 2
		Experiment 1 ($N = 205$)	
Adults	Tone effect	Silence	Speech effect
		$R^2 = .11, p < .01$	$\Delta R^2 = .28, p < .01$
Adults	Speech effect	Silence	Tone effect
		$R^2 = .06, p < .01$	$\Delta R^2 = .29, p < .01$
		Experiment 2 ($N = 64$ adults, 63 children)	
Adults	Tone effect	Silence	Speech effect
		$R^2 = .17, p < .01$	$\Delta R^2 = .11, p < .01$
Adults	Speech effect	Silence	Tone effect
		$R^2 = .12, p < .01$	$\Delta R^2 = .12, p < .01$
Children	Tone effect	Silence	Speech effect
		$R^2 = .25, p < .001$	$\Delta R^2 = .14, p < .001$
Children	Speech effect	Silence	Tone effect
		$R^2 = .10, p < .05$	$\Delta R^2 = .16, p < .001$

Note—The data for Experiment 2 included both span and span $-$ 1 list lengths averaged together.
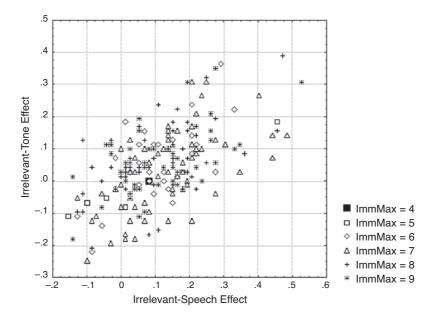
Figure 1. Scatterplot of the irrelevant-tone effect as a function of the irrelevant-speech effect in Experiment 1 (in proportion difference scores). Separate symbols (graph parameter) represent individuals with different immediate maximal span (Imm-Max) scores.

One possible reason for this absence of correlation in the adult samples in which it has been evaluated is that avoidance of the ISE requires attention (Buchner et al., 2004; Cowan, 1995) and therefore would correlate more highly with WM procedures in which processing and storage are combined (e.g., reading span, Daneman & Carpenter, 1980; operation span, Turner & Engle, 1989), given that these tasks seem to depend on how well attention can be controlled (e.g., Kane & Engle, 2003). Although this hypothesis was not confirmed in one recent study with adult participants (Beaman, 2004), in Experi-ment 2 we reexamined this issue with samples of second-grade children and adults, using both irrelevant speech and irrelevant tones, as in Experiment 1. In this experiment, the list length depended on the individual's span, so that the level of difficulty could be kept comparable across age groups.

One reason to examine children, as well as adults, is that children do not engage in systematic, effective covert verbal rehearsal, as adults do (e.g., Cowan & Kail, 1996; Flavell, Beach, & Chinsky, 1966; Ornstein & Naus, 1978). The inclusion of children in Experiment 2, with rapidly

Table 4
Regression Analyses in Experiments 1 and 2 Using Condition Scores

Age Group	Dependent Measure	Regression Step 1	Regression Step 2
		Experiment 1 ($N = 205$)	
Adults	Tones	Silence	Speech
		$R^2 = .61, p < .01$	$\Delta R^2 = .11, p < .01$
Adults	Speech	Silence	Tones
		$R^2 = .56, p < .01$	$\Delta R^2 = .13, p < .01$
		Experiment 2 ($N = 64$ adults, 63 children)	
Adults	Tones	Silence	Speech
		$R^2 = .22, p < .01$	$\Delta R^2 = .11, p < .01$
Adults	Speech	Silence	Tones
		$R^2 = .19, p < .01$	$\Delta R^2 = .11, p < .01$
Children	Tones	Silence	Speech
		$R^2 = .36, p < .001$	$\Delta R^2 = .12, p < .001$
Children	Speech	Silence	Tones
		$R^2 = .26, p < .001$	$\Delta R^2 = .13, p < .001$

Note—The data for Experiment 2 included both span and span − 1 list lengths averaged together.

Table 5
Regression Analyses in Experiment 1
Using Span Scores as the Dependent Measure

Dependent Measure	Regression Step 1	Regression Step 2	Regression Step 3
DelCum	Silence $R^2 = .52, p < .001$	Speech $\Delta R^2 = .02, p < .01$ Tones $\Delta R^2 = .02, p < .01$	Tones $\Delta R^2 = .01, p < .05$ Speech $\Delta R^2 = .00$, n.s.
DelMax	Silence $R^2 = .44, p < .001$	Speech $\Delta R^2 = .00$, n.s. Tones $\Delta R^2 = .00$, n.s.	Tones $\Delta R^2 = .01$, n.s. Speech $\Delta R^2 = .00$, n.s.
ImmCum	Silence $R^2 = .44, p < .001$	Speech $\Delta R^2 = .03, p < .01$ Tones $\Delta R^2 = .03, p < .001$	Tones $\Delta R^2 = .01, p < .05$ Speech $\Delta R^2 = .00$, n.s.
ImmMax	Silence $R^2 = .32, p < .001$	Speech $\Delta R^2 = .02, p < .05$ Tones $\Delta R^2 = .01$, n.s.	Tones $\Delta R^2 = .00$, n.s. Speech $\Delta R^2 = .01, p < .05$

Note—Span task measures: DelCum, delayed test, cumulative score; DelMax, delayed test, maximum score; ImmCum, immediate test, cumulative score; ImmMax, immediate test, maximum score.

changing abilities of not only rehearsal but also attention (Lane & Pearson, 1982; Zukier & Hagen, 1978), helped us address the issue of correlations between measures of span and the ISE by including more individual difference variability. Furthermore, we already know that there is an increase in the ISE with speech as opposed to tone distractors in children, more than in adults (Elliott, 2002), but the relation between irrelevant-speech and irrelevant-tone effects has not been examined in this age group.

EXPERIMENT 2
Various Span Measures and the ISE

Method

Participants

Sixty-four adults (43 female, 21 male) from the University of Missouri and 63 children (34 female, 29 male; ages 7 years, 1 month to 9 years, 2 months; $M = 8$ years, 2 months, $SD = 4$ months) in the 2nd grade from the Columbia public school system were included in these analyses. Adults were given credit in their psychology classes, and the children were given $5 and a book for their participation. All participants reported normal or corrected-to-normal vision and normal hearing, and were native speakers of English.

Design and Procedure

The participants were asked to complete a visual span task, an auditory span task, a reading span task, an operation span task, and an irrelevant-sound task that included both speech and tones. The visual span task was identical to the immediate span task used in Experiment 1, and the auditory span task was only different in that the digits to be recalled were presented over headphones.

The two complex span tasks were slight variations of the reading and operation span tasks used by Towse, Hitch, and Hutton (1998). In the reading span task, sentences were presented that were missing the final word. The participants were to read aloud each sentence, complete it with a single word, and recall these generated words when a cue was presented (e.g., "The clown had a big smiling ___;" the expected answer was "face"). Each trial included two to seven sentences (randomly varying from 5 to 13 words long, including the completion word), with three trials at each set size. If a

participant correctly repeated all of the completions for at least one trial, the computer advanced to the next set size.

The stimuli for the operation span task were arithmetic problems presented in the form of A [op] B = ? (short form) or A [op] B [op] C [op] D, where [op] was "+" or "−," A was a number from 1 to 10, and B through D were each 0 or 1. The sum always fell in the range of 3 to 9—for example, "4 + 1 − 1 + 1 = ?" The participant read aloud the arithmetic problem, solved the problem, and tried to retain the sum. These problems were presented in set sizes ranging from two to seven problems. The participant verbally recalled the sums in order when a cue to recall was presented. Two lists had to be recalled correctly within a set size for the computer to advance to the next set size. If a participant recalled two consecutive lists correctly, the program advanced to the next set size without presenting the third problem for that set size.

To control for the level of difficulty across the age groups, lists in the irrelevant-sound task were presented at each individual's measure of maximum visual span and at span minus one (span − 1). Further details on the method of the irrelevant-sound task were reported in Elliott (2002).

Results and Discussion

Means

The means and standard errors for the span measures and the irrelevant-sound task are shown in Table 6. As in Experiment 1, difference scores were created to assess the speech and tone effects.

Correlations

The correlations between the measures are shown in Table 7 for the adults and Table 8 for the children. In the adults, there was a significant correlation between the irrelevant-speech effect and maximal visual span, in the unanticipated direction (i.e., higher spans corresponding to a larger irrelevant-speech effect; $r = .38, p = .002$). After correcting for attenuation of the measures, this was a very large correlation ($r = .95$). We did a power analysis for this adult correlation, and with $N = 64$ (and an es-

Table 6
Means (and Standard Errors) in Experiment 2

	Age Group					
	Children			Adults		
Measure	N	M	SEM	N	M	SEM
Span task						
VspanMax	63	5.0	0.10	64	7.3	0.15
VspanCum	63	4.3	0.09	64	6.6	0.14
AspanMax	59	5.6	0.11	64	7.5	0.13
AspanCum	59	4.9	0.09	64	6.7	0.12
RspanMax	60	2.5	0.09	64	4.6	0.12
RspanCum	60	2.0	0.06	64	3.8	0.09
OspanMax	55	3.6	0.13	64	6.1	0.11
OspanCum	55	3.1	0.11	64	5.5	0.12
List-recall task						
Span length lists						
Silence	63	0.77[a]	0.03	64	0.79[a]	0.02
Speech	63	0.41[b]	0.03	64	0.70[b]	0.02
Tones	63	0.56[c]	0.03	64	0.71[b]	0.02
Span − 1 length lists						
Silence	63	0.89[a]	0.02	64	0.87[a]	0.02
Speech	63	0.64[b]	0.03	64	0.82[b]	0.02
Tones	63	0.80[c]	0.03	64	0.90[a]	0.02
Average of both list lengths						
Silence	63	0.83[a]	0.02	64	0.83[a]	0.01
Speech	63	0.52[b]	0.03	64	0.76[b]	0.02
Tones	63	0.68[c]	0.02	64	0.80[a]	0.02

Note—Span test measures: Vspan, visual digit span; Aspan, auditory digit span; Rspan, reading span; Ospan, operation span; Max, maximum span measure; Cum, cumulative span measure. Within each age group, list-recall condition scores marked with different superscript letters (a–c) are significantly different in Newman–Keuls post hoc tests.

timate of $\rho = .38$), the power to detect this correlation was .88. With this sample, $r = .34$ could be detected with .80 power. (In children, no such correlation was significant [$r = -.03$]. In our sample of 63 children, to detect a correlation with a minimum value of .35, the power was .81.)

One possible explanation for this correlation between the irrelevant-speech effect and maximal visual span in adults is that irrelevant speech may interfere with rehearsal, as many theorists agree (Beaman & Jones, 1997, 1998; Jones, 1993). The explanation would be as follows. If one examines the number of items recalled in the silent control condition, it increases markedly with span size. For individuals with spans (and list lengths in the ISE procedure) of 5, 6, 7, 8, and 9, respectively ($n = 3$, 16, 17, 13, and 15, respectively), the mean numbers of items correct were 4.44, 5.04, 5.31, 6.12, and 7.15. In contrast, in the irrelevant-speech condition, the numbers

Table 7
Correlations Between Measures in Adult Participants in Experiment 2

	1	2	3	4	5	6	7	8	9	10	11	12	13
VspanMax (1)	.85	1.00	.74	.75	.34	.57	.49	.60	−.36	−.77	−.25	.95	−.11
VspanCum (2)	.91*	.86	.72	.77	.32	.54	.45	.57	−.25	−.51	.00	.62	−.37
AspanMax (3)	.62*	.60*	.82	1.00	.48	.66	.52	.61	.05	−.19	.06	.41	−.05
AspanCum (4)	.61*	.63*	.86*	.78	.53	.74	.48	.65	.23	−.16	.18	.65	.06
RspanMax (5)	.26*	.25*	.36*	.39*	.69	1.00	.61	.53	.35	.25	.38	−.03	−.15
RspanCum (6)	.42*	.40*	.48*	.52*	.82*	.64	.72	.77	.22	.07	.09	.17	.18
OspanMax (7)	.40*	.37*	.42*	.38*	.45*	.51*	1.00	.10	−.25	−.08	.57	.30
OspanCum (8)	.50*	.48*	.50*	.52*	.40*	.56*	.86*	.82	−.06	−.35	−.10	.56	.08
Silence (9)	−.23	−.16	.03	.14	.20	.12	.06	−.04	.47	.84	.99	1.00	1.00
Speech (10)	−.54*	−.36*	−.13	−.11	.16	.04	−.17	−.24	.44*	.58	.86	−1.00	−.38
Tones (11)	−.16	.00	.04	.11	.22	.05	−.05	−.06	.47*	.50*	.48	−.50	−1.00
SpeechEffect (12)	.38*	.25	.16	.25	−.01	.06	.22	.22	.35*	−.69*	−.15	.19	1.00
ToneEffect (13)	−.04	−.14	−.02	.02	−.05	.06	.11	.03	.41*	−.12	−.61*	.46*	.17

Note—Vspan, visual digit span; Aspan, auditory digit span; Rspan, reading span; Ospan, operation span; Max, maximum span measure; Cum, cumulative span measure. $N = 64$. Data points were averaged across both span and span −1 length lists. Scores on the diagonal (in italic) represent Cronbach's alpha reliability coefficients, with the exception of the scores for the Ospan and Vspan tasks. For those tasks, Spearman–Brown split-half reliability estimates were calculated. Scores above the diagonal represent the correlations after correction for attenuation. These were not evaluated for statistical significance.   *$p < .05$.

Table 8
Correlations Between Measures in Child Participants in Experiment 2

	1	2	3	4	5	6	7	8	9	10	11	12	13
VspanMax (1)	*.66*	1.00	.50	.59	.46	.59	.36	.24	−.74	−.52	−.86	−.06	0
VspanCum (2)	.77*	*.83*	.52	.61	.20	.51	.35	.20	−.40	−.36	−.54	.11	0
AspanMax (3)	.34*	.39*	*.70*	1.00	.20	.36	.23	.31	.23	.37	.08	−.32	0
AspanCum (4)	.39*	.45*	.79*	*.65*	.36	.50	.27	.35	.10	.10	−.12	−.06	0
RspanMax (5)	.22	.11	.10	.17	*.35*	1.00	.68	.40	−.18	.22	.22	−.53	0
RspanCum (6)	.30*	.29*	.19	.25	.69*	*.39*	.29	.31	−.27	.04	−.20	−.34	0
OspanMax (7)	.26	.28*	.16	.19	.16	.16	*.78*	1.00	.01	−.37	−.13	.56	0
OspanCum (8)	.17	.16	.23	.25	.21	.18	.89*	*.77*	.08	−.23	−.07	.43	0
Silence (9)	−.50*	−.30*	.16	.07	−.09	−.14	.01	.06	*.69*	.70	1.00	.55	0
Speech (10)	−.35*	−.27	.26	.07	.11	.02	−.27	−.17	.48*	*.69*	.88	−1.00	0
Tones (11)	−.44*	−.31*	.04	−.06	.08	−.08	−.07	−.04	.56*	.61*	*.39*	−.53	0
SpeechEffect (12)	−.03	.06	−.16	−.03	−.19	−.13	.30*	.23	.28*	−.71*	−.21	*.37*	0
ToneEffect (13)	−.09	.00	.13	.14	−.18	−.07	.09	.11	.51*	−.11	−.43*	.53*	*0*

Note—Vspan, visual digit span; Aspan, auditory digit span; Rspan, reading span; Ospan, operation span; Max, maximum span measure; Cum, cumulative span measure. Based on $N = 52$ participants with data for all of the measures. Data points were averaged across both span and span − 1 length lists. Scores on the diagonal (in italic) represent Cronbach's alpha reliability coefficients, with the exception of the scores for the Ospan and Vspan tasks. For those tasks, Spearman–Brown split-half reliability estimates were calculated. Scores above the diagonal represent the correlations after correction for attenuation. These were not evaluated for statistical significance.   *$p < .05$.

of items correct in these subgroups were much more nearly identical: 4.56, 4.55, 5.15, 5.12, and 5.39. Thus, irrelevant speech greatly diminished the ability of individuals with a longer span to correctly recall the longer lists that they received in this experiment. Given that the ISE was calculated as a difference between the silent and speech conditions, this would have resulted in larger ISEs for more capable participants.

In the children, there was a finding of a significant correlation between the maximum measure of operation span (OspanMax) and the irrelevant-speech effect difference score ($r = .30$), as is shown in Table 8. When this correlation was adjusted for attenuation, it increased to $r = .56$. This may suggest, as discussed above, that children with more sophisticated rehearsal processes were more susceptible to the ISE. However, no other correlations between measures of simple or complex span and the speech effect were significant in the children, so it is difficult to clearly interpret the correlation with OspanMax. (As mentioned above, for the power to detect a significant correlation to reach the acceptable level of .80, the value of r needed to be at least .35.) Regardless of the exact mechanism underlying this correlation, the most important point is that it was not in the expected direction. The higher span individuals had larger irrelevant-speech effects, and that requires a revised theoretical analysis, which we will offer in the General Discussion section.

Regressions

We conducted the same regression analyses on the data separately for the adults and the children. For the adults, the results were comparable to those obtained in Experiment 1. In accounting for the tone effect, after variation due to the silent control condition was removed, there was considerable residual variance due to the speech effect. Conversely, in accounting for the speech effect, after variation due to the silent control condition was removed, there was considerable residual variance due to

the tone effect. This same pattern was also seen in the children, indicating that these effects have shared variance in both the adults and the children (see Table 3). Analyses in which the condition means were used instead of the difference scores also produced the same significant effects as in Experiment 1, for both adults and children (see Table 4).

As with Experiment 1, another series of regression analyses were conducted using the measures of memory span (simple and complex measures in this experiment) as the dependent variables to determine whether the paucity of significant relationships between our span measures and the effect scores was due to the unreliability of the effect scores. We found that this was not the case, since the analyses using raw measures roughly paralleled our findings with the effect scores. When we averaged the two list lengths, the only regression analyses to reach significance in the adults were when the immediate visual span measures were used (referred to as VspanMax and VspanCum to differentiate these from the auditory span measures). When VspanCum was used as the dependent measure, after silence was taken out in the first step of the analyses, for the speech condition $\Delta R^2 = .10$, $p < .01$, and the variance explained by the tone condition was not significant. When done in the reverse order, the tone condition did not explain significant variance but the speech condition resulted in $\Delta R^2 = .14$, $p < .01$. When VspanMax was used as the dependent measure, the tone condition once more did not explain a significant amount of variance in either order. The speech condition accounted for $\Delta R^2 = .24$, $p < .01$ when it came before the tone condition, and $\Delta R^2 = .25$, $p < .01$ when following the tone condition. Notice from the raw correlation (in Table 7) that individuals with higher spans actually had lower performance levels in the irrelevant-speech condition.

The same analyses were conducted with the children's data. In line with the correlational analyses, the only significant finding in the regressions was for OspanMax.

The silent control condition accounted for no variance in OspanMax, and the irrelevant-speech condition, entered subsequently, accounted for $\Delta R^2 = .10$, $p < .05$. Again, as in the visual spans in adults, notice from the raw correlation (in Table 8) that children with higher OspanMax scores actually had lower performance levels in the irrelevant-speech condition.

It appears that the only way to account for this pattern of findings, which includes larger ISEs for higher span individuals even after the silent-condition variance is removed, is by an appeal to individual differences in strategies or processing styles. For example, it may be that using rehearsal is a good strategy for span performance but a bad strategy for remembering a printed list in the presence of speech interference; perhaps, in the presence of speech interference, rehearsal is worse than some alternative strategy such as simply trying to focus attention on the items to be remembered (cf. Cowan, 2001). If individuals tend not to be aware of the need to switch strategies in the presence of irrelevant speech, that could result in poorer performance in the irrelevant-speech condition in individuals with higher spans. Yet it is not clear why these correlations were not found more consistently.

GENERAL DISCUSSION

This study is the first to use a large sample and regression to investigate individual differences in the disruption of serial recall by irrelevant sounds. We now have evidence that the degree of susceptibility to the ISE is a reliable individual trait that has a substantial materials-general component, as opposed to materials specific. The results of these experiments indicate a strong relation between irrelevant-tone and irrelevant-speech effects, independent of span level.

For the most part, we have confirmed the previous (Beaman, 2004; Ellermeier & Zimmer, 1997; Neath et al., 2003) indications that this susceptibility to the ISE is unrelated to memory span. However, we have also demonstrated the existence of various levels of susceptibility to the ISE, consistent across speech and tone effects, for participants at *each* level of span (see Figure 1). Thus, clearly, the absence of a correlation between the ISE and span cannot be attributed to unreliability of the ISE.

In several instances, we found positive correlations between the magnitude of the ISE and different types of memory span. In Experiment 1, the delayed maximum span showed this effect, for both irrelevant-speech and irrelevant-tone effects. In Experiment 2, the visual maximum span for adults and the operation span for children showed it. All of these correlations run contrary to the a priori expectation of a *smaller* ISE in more capable individuals.

In a post hoc manner, these positive correlations can be explained in several ways. First, if higher span individuals excel because they carry out mnemonic processing such as rehearsal more efficiently than do lower span individuals, and if this mnemonic processing is interrupted by irrelevant sounds (e.g., Beaman & Jones, 1997, 1998; Jones, 1993), larger ISEs would be expected in individuals who did more of the mnemonic processing in the first place.

Second, though, perhaps the correlation between high spans and the ISE is a statistical artifact in that spans are related to the silent condition used to calculate the ISE. However, this cannot be the sole explanation, given the regressions showing that span variance was picked up by the irrelevant-speech and irrelevant-tone conditions even after the variance from the silent control condition was removed. To explain these regression results, an account like the one involving mnemonic processing and its interruption by the ISE appears to be warranted.

These results provide one possible explanation for the puzzling absence of *negative* correlations between various sorts of working-memory span and the ISE (Beaman, 2004; Ellermeier & Zimmer, 1997; Neath, Farley, & Surprenant, 2003). A mechanism that, by itself, leads toward such negative correlations theoretically may still exist, but it may be balanced out (or sometimes outweighed) by the factors leading toward a positive correlation, as was discussed above.

The second experiment of this study supported the findings of the first and also included a developmental comparison. Given the dramatic decrease in the magnitude of the ISEs with development in childhood (Elliott, 2002), along with the dramatic increase in the magnitude of span itself, it is still surprising that negative correlations between span and the ISE have not been observed within an age group. This is surprising because one might have imagined the same mechanisms that account for development of span to account for the developmental decrease in the magnitude of the ISE. Such factors could include the developmental improvement in both covert verbal rehearsal ability (e.g., Cowan & Kail, 1996; Flavell et al., 1966; Ornstein & Naus, 1978) and the control of attention (e.g., Lane & Pearson, 1982; Zukier & Hagen, 1978). It seemed natural to suppose that the same factors would operate within an age group to cause individuals with higher spans to have smaller ISEs.

An alternative possibility reinforced by the evidence is that there are rather different influences on span and the magnitude of the ISE. For example, whereas a more sophisticated strategy of covert verbal rehearsal undoubtedly is helpful to span in a silent setting, in the presence of irrelevant speech there may be both advantages and drawbacks to such a mnemonic strategy. It is also possible that the ability to control attention is more important for overcoming the ISE than for span in a silent setting.

There are many possible ways to explain the interesting discrepancy between the developmental results (Elliott, 2002, and the present study) and the individual-difference results of the present study. We offer just one scenario as an illustration. Within an individual, rehearsal may be a good strategy for span tasks but a poor strategy to overcome ISEs (at least in the case of speech interference). Across age groups, however, the adults, who are better able to rehearse, also have more advanced attentional ca-

pabilities, and that could be the predominant effect leading to the developmental decrease in the ISE.

Let us consider the contribution of attention in somewhat more detail, since its role in processing may be complex. Previous evidence from the cross-modal Stroop paradigm, in which participants are asked to name a colored square as quickly as possible while ignoring spoken, incongruent color words (Cowan & Barron, 1987; Elliott, Cowan, & Valle-Inclán, 1998), supports the conjecture that attention is helpful to overcome at least some sorts of ISEs. When examined developmentally, the magnitude of the disruption in the cross-modal Stroop effect was shown to decrease with age (Hanauer & Brooks, 2003), which the authors interpreted in terms of developmental improvements in selective attention. In contrast, there is little evidence that the control of attention is a key factor in verbal span tasks, although it may be more important for tasks in which rehearsal is prevented (Cowan, 2001).

If the control of attention is important for overcoming the ISE, one would have expected it to correlate with performance in WM tasks in which the control of attention is important. One such WM task is operation span. Conway, Cowan, and Bunting (2001) demonstrated that the ability to overcome auditory distraction is related to operation span. They examined individuals in the upper and lower quartiles of a larger sample of participants who completed an operation span task. Then, using a dichotic listening procedure, high- and low-WM-span participants shadowed (or repeated immediately after hearing) words presented to one ear, while ignoring words presented to the other ear. After several minutes of shadowing, the participant's name was presented in the unattended channel. The results were striking: Only 20% of high-span participants reported hearing their name in the unattended channel, in contrast to 65% of low-span participants. Thus, in adults, WM capacity has been shown to be an important component of avoiding distraction in a dichotic listening task. The puzzle is, then, the fact that, in our Experiment 2, operation span was not negatively related to the magnitude of the ISE.

The solution to this puzzle, to which we have hinted above, is that another contaminating factor heavily influences the magnitude of the ISE. Individuals who have good control of attention may also tend to adopt a covert verbal rehearsal strategy, but that strategy may be harmful when irrelevant speech is present. Some positive correlations between the magnitude of the ISE and some span measures in the present study indicate why it would be difficult to observe a theoretically expected negative correlation, even if the hypothetical mechanism underlying such a correlation (such as the ability to ignore irrelevant sounds) does, in fact, exist.

In conclusion, the present research contributes to the literature on the ISE by (1) confirming previous findings regarding the absence of a general relationship between memory span and the ISE and (2) demonstrating a significant amount of shared variance between speech and

tone effects in adults and in children. The latter finding suggests a role for a common mechanism of disruption, regardless of the nature of the irrelevant sounds. The relations between WM spans and the ISE that were sporadically obtained were in the unexpected direction (larger ISEs for more capable participants), suggesting that a strategy often used to good effect in span tasks, such as covert verbal rehearsal, may actually be counterproductive in the presence of irrelevant sounds (at least irrelevant speech). The exact nature of the mechanisms involved, and the exact roles of attention, rehearsal, and episodic encoding in the ISE, are key questions for future research.

REFERENCES

BADDELEY, A. D. (2000). The phonological loop and the irrelevant speech effect: Some comments on Neath (2000). *Psychonomic Bulletin & Review*, **7**, 544-549.

BEAMAN, C. P. (2004). The irrelevant sound phenomenon revisited: What role for working memory capacity? *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **30**, 1106-1118.

BEAMAN, C. P., & JONES, D. M. (1997). Role of serial order in the irrelevant speech effect: Tests of the changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 459-471.

BEAMAN, C. P., & JONES, D. M. (1998). Irrelevant sound disrupts order information in free recall as in serial recall. *Quarterly Journal of Experimental Psychology*, **51A**, 615-636.

BUCHNER, A., ROTHERMUND, K., WENTURA, D., & MEHL, B. (2004). Valence of distractor words increases the effects of irrelevant speech on serial recall. *Memory & Cognition*, **32**, 722-731.

COLLE, H. A., & WELSH, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning & Verbal Behavior*, **15**, 17-31.

CONWAY, A. R. A., COWAN, N., & BUNTING, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, **8**, 331-335.

COWAN, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.

COWAN, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral & Brain Sciences*, **24**, 87-185.

COWAN, N., & BARRON, A. (1987). Cross-modal, auditory–visual Stroop interference and possible implications for speech memory. *Perception & Psychophysics*, **41**, 393-401.

COWAN, N., & KAIL, R. (1996). Covert processes and their development in short-term memory. In S. Gathercole (Ed.), *Models of short-term memory* (pp. 29-50). Hove, U.K.: Erlbaum.

DANEMAN, M., & CARPENTER, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, **19**, 450-466.

ELLERMEIER, W., & ZIMMER, K. (1997). Individual differences in susceptibility to the "irrelevant speech effect." *Journal of the Acoustical Society of America*, **102**, 2191-2199.

ELLIOTT, E. M. (2001). *Developmental differences in the effects of distracting sounds on performance*. Unpublished doctoral dissertation, University of Missouri, Columbia.

ELLIOTT, E. M. (2002). The irrelevant-speech effect and children: Theoretical implications of developmental change. *Memory & Cognition*, **30**, 478-487.

ELLIOTT, E. M., COWAN, N., & VALLE-INCLÁN, F. (1998). The nature of cross-modal, color–word interference effects. *Perception & Psychophysics*, **60**, 761-767.

FLAVELL, J. H., BEACH, D. R., & CHINSKY, J. M. (1966). Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development*, **37**, 283-299.

HANAUER, J. B., & BROOKS, P. J. (2003). Developmental change in the cross-modal Stroop effect. *Perception & Psychophysics*, **65**, 359-366.

JONES, D. M. (1993). Objects, streams, and threads of auditory attention. In A. D. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, aware-*

ness, and control. A tribute to Donald Broadbent (pp. 87-104). Oxford: Oxford University Press.

JONES, D. M., & MACKEN, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 369-381.

JONES, D. M., & TREMBLAY, S. (2000). Interference in memory by process or content? A reply to Neath (2000). *Psychonomic Bulletin & Review*, **7**, 550-558.

KANE, M. J., & ENGLE, R. W. (2003). Working memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, **132**, 47-70.

LANE, D. M., & PEARSON, D. A. (1982). The development of selective attention. *Merrill-Palmer Quarterly*, **28**, 317-337.

NEATH, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review*, **7**, 403-423.

NEATH, I., FARLEY, L. A., & SURPRENANT, A. M. (2003). Directly assessing the relationship between irrelevant speech and articulatory suppression. *Quarterly Journal of Experimental Psychology*, **56**, 1269-1278.

ORNSTEIN, P. A., & NAUS, M. J. (1978). Rehearsal processes in children's memory. In P. A. Ornstein (Ed.), *Memory development in children* (pp. 69-99). Hillsdale, NJ: Erlbaum.

SALAMÉ, P., & BADDELEY, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning & Verbal Behavior*, **21**, 150-164.

TOWSE, J. N., HITCH, G. J., & HUTTON, U. (1998). A reevaluation of working memory capacity in children. *Journal of Memory & Language*, **39**, 195-217.

TREMBLAY, S., & JONES, D. M. (1998). Role of habituation in the irrelevant sound effect: Evidence from the effects of token set size and rate of transition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 659-671.

TURNER, M. L., & ENGLE, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory & Language*, **28**, 127-154.

ZUKIER, H., & HAGEN, J. W. (1978). The development of selective attention under distracting conditions. *Child Development*, **49**, 870-873.