Perception & Psychophysics
2007, 69 (5), 757-771

Task-dependent costs in processing two simultaneous auditory stimuli

Frederick J. Gallun, Christine R. Mason, and Gerald Kidd, Jr.
Boston University, Boston, Massachusetts

A listener presented with two speech signals must at times sacrifice the processing of one signal in order to understand the other. This study was designed to distinguish costs related to interference from a second signal (*selective attention*) from costs related to performing two tasks simultaneously (*divided attention*). Listeners presented with two processed speech-in-noise stimuli, one to each ear, either (1) *identified* keywords in both or (2) *identified* keywords in one and *detected* the presence of speech in the other. Listeners either knew which ear to report in advance (single task) or were cued afterward (partial-report dual task). When the dual task required two identification judgments, performance suffered relative to the single-task condition (as measured by percent correct judgments). Two different tasks (identification for one stimulus and detection for the other) resulted in much smaller reductions in performance when the cue came afterward. We concluded that the degree to which listeners can simultaneously process dichotic speech stimuli seems to depend not only on the amount of interference between the two stimuli, but also on whether there is competition for limited processing resources. We suggest several specific hypotheses as to the structural mechanisms that could constitute these limited resources.

When presented with two speech utterances, one to each ear (i.e., *dichotically*), listeners can be asked to either (1) ignore one and report the other (*selective attention*) or (2) report both (*divided attention*). Despite more than 50 years of research on this topic, it is still not fully understood why different stimulus configurations exert such a great influence on the degree of success listeners experience when asked to either select one utterance or divide their auditory attention between two utterances. The goal of this study was to explore the theoretical distinction between selective and divided attention by presenting a method for distinguishing between the two processes in a single study. The new set of experimental results demonstrates the value of measuring both processes within the same study and provides an example of the sorts of hypotheses that such a technique can be used to generate.

Two different types of experimental approaches have generally been used to examine the processing of multiple speech stimuli. The first type is what we will refer to as the *dual-ear* experiment (although, for comparisons, most such experiments have contained single-ear and/or diotic conditions as well). Dual-ear experiments are distinguished from other experiments involving speech stimuli by the fact that they include the presentation of one speech utterance to one ear and another utterance to the other ear, and in terms of results, they tend to emphasize the fact that under some conditions listeners can reliably report the information presented to the "target" ear with very few errors (e.g., Broadbent, 1958; Cherry, 1953; Moray, 1970; Treisman, 1964, 1969; Wood & Cowan, 1995).

The second type is the *informational masking* approach, in which researchers have generally presented multiple speech stimuli to the *same* ear (although various types of dichotic presentation have also been used) and have emphasized the factors that lead to *errors* in processing only one of two simultaneously presented stimuli (e.g., Brungart, Simpson, Ericson, & Scott, 2001). The term *informational masking* is often defined in contrast to *energetic masking*, which refers to a reduction in performance that can be accounted for by the degree to which the masker overlaps the target at a set of peripheral analyzers (usually the cochlea or the auditory nerve). Informational masking, it is argued, is caused by nonperipheral factors, principally masker uncertainty and target–masker similarity (for examples, see Durlach et al., 2003; Neff, 1995; Neff & Green, 1987; Watson & Kelly, 1981; for a review, see Kidd, Mason, Richards, Gallun, & Durlach, 2006). A substantial number of studies have demonstrated that speech stimuli can interfere with each other in ways that are not easily captured by energetic masking alone (e.g., Arbogast, Mason, & Kidd, 2002; Brungart & Simpson, 2002; Brungart, Simpson, Darwin, Arbogast, & Kidd, 2005; Brungart et al., 2001; Freyman, Balakrishnan, & Helfer, 2001, 2004; Freyman, Helfer, McCall, & Clifton, 1999; Gallun, Mason, & Kidd, 2005; Kidd, Arbogast, Mason, & Gallun, 2005; Kidd, Mason, & Gallun, 2005; Shinn-Cunningham & Ihlefeld, 2004). Although the roles of similarity and uncertainty have been less clearly articulated for speech stimuli, for the purpose of this discussion the term *informational masking* will be primarily used in reference to speech studies.

F. J. Gallun, frederick.gallun@va.gov

Although the two experimental approaches (dual-ear and informational masking) emphasize different aspects of how multiple speech stimuli are processed by human listeners and often use different stimulus configurations, their findings are generally compatible. Cherry (1953), for example, reported that listeners asked to repeat the words spoken to one ear (a "shadowing" task) were often unaware of whether the language being spoken in the other ear had changed or even if it had been switched to being played backward. On the other hand, they were able to detect a change in the speaker's gender or a switch from speech to a tone in the unattended ear (Cherry, 1953). Since the stimuli at the two ears fall into well-separated peripheral analyzers, no interference is predicted by peripheral overlap alone, making some of Cherry's results quite similar to demonstrations of informational masking. Unlike most studies of informational masking, however, many of these results have been shown to depend on a specific central processing resource: memory. Wood and Cowan (1995) showed that the length of shadowing and the time at which changes were inserted influenced the detectability of these changes, suggesting that listeners were aware of the unattended stimulus but did not retain it in memory.

In addition, Wood and Cowan (1995) found that listeners who detected switches in the nontarget speech from forward to backward presentation or who recalled words spoken in the nontarget ear made errors in the primary shadowing task. This suggests that a listener who dutifully attends to only one ear is unlikely to detect more than gross changes to the long-term statistics of the unattended ear (e.g., a low-pitched voice changing to a high-pitched one or speech changing to a tone.) Finally, it is important to note that listeners in dual-ear tasks were not informed in advance that they would be asked to make responses based on the unattended channel. When listeners are prepared to make two responses, there are reports suggesting that performance can be as good in a divided-attention task as in a selective-attention task (Moore & Massaro, 1973; Shiffrin, Pisoni, & Castaneda-Mendez, 1974)

The role of memory as a structural constraint governing the processing of multiple stimuli has yet to be investigated to an extent that would allow it to be included in a computational model of such processing, and this constraint is generally not considered in studies of informational masking. There is, however, work from dual-sensory studies that points in a similar direction. Hafter and his colleagues (Bonnel & Hafter, 1998; Hafter, Bonnel, Gallun, & Cohen, 1998) explained a difference that they observed between detection and identification in a dual-sensory dual-task study by positing a role for memory in the ability to divide attention. They hypothesized that stimulus comparisons relying on what Durlach and Braida (1969) called "sensory-trace" memory do not interfere across modalities, but that comparisons involving "context-coding" (Durlach & Braida's other type of memory for sensory events) do result in cross-modal interference. This distinction suggests that structural constraints can be thought of as involving the inability of specialized neural structures to perform operations in parallel. From this perspective, dual-task costs indicate that either (1) the necessary information cannot be extracted by multiple structures simultaneously or (2) the unprocessed input cannot be held in memory until the processor is available. The success of this approach recommends its application to studies of multiple speech stimuli as well.

Although her concern was not with memory, Treisman (1969) also invoked structural constraints in the central processing system in order to explain interference from an irrelevant speech source in a shadowing task. In her framework, interference results from competition for "processing resources" at the level of *feature analyzers* rather than from direct competition at the level of input. She hypothesized that both speech inputs are processed in parallel to some level of analysis, but that only one can be analyzed by the speech recognition system at one time. Presumably, any cues that listeners can use to distinguish which input should be selected for further analysis should improve performance. This is in agreement with the finding from informational masking studies that, when available, listeners use perceived location cues to help them choose the correct utterance to process (Brungart et al., 2005; Freyman et al., 2004; Gallun et al., 2005; Kidd, Arbogast, et al., 2005; Shinn-Cunningham & Ihlefeld, 2004). Similarly, Brungart et al. (2001) found that knowing the target talker's voice in advance also served as a cue that listeners could use to improve performance. Treisman's (1969) framework also explains why some features of the non-shadowed stimuli can be recognized while others cannot (Cherry, 1953; Wood & Cowan, 1995). The pitch of a voice or the difference between a tone and speech may be a "low-level" feature that is extracted before the speech recognition system even gets involved.

Norman and Bobrow (1975), Navon and Gopher (1979), and Wickens (1984) proposed a general approach to competition for central processing resources when they suggested that the human perceptual system is composed of multiple types of processing resources and that whether or not interference is observed depends on the degree to which multiple tasks require access to the same resources. In this conceptualization, *processing resources* refers to everything from input channels or feature analyzers to system processes like the retrieval of representations from short- or long-term memory.

DESIGN OVERVIEW

Costs in an identification task were compared with the costs in a task involving the detection of speech in noise. When the two tasks were the same (both identification), good performance would require concurrent access to the same processing resources, and thus it was hypothesized that there would be a cost associated with performing both tasks. Alternatively, we predicted no dual-task cost for the situation in which the tasks were different. Although our setup parallels the experiments of Bonnel and Hafter (1998), the use of within-modality rather than between-modality interference (as well as the use of speech) suggests that the results might be different.

The stimuli were filtered into mutually exclusive non-overlapping frequency bands (Arbogast et al., 2002) in order to allow the frequencies of the two sentences to be chosen randomly on each trial and to remove pitch characteristics from the speakers' voices. By ensuring that listeners would be required to make use of the ear of presentation to distinguish the two talkers, the number of possible strategies listeners could employ was reduced. Frequency randomization has also been shown to be associated with informational masking for nonspeech stimuli (see, e.g., Lutfi, 1993), so it was anticipated that selecting new frequency bands on every trial would increase the interference observed. Nonoverlapping bands were used in order to eliminate the across-ear interactions that can occur when a small number of the same narrow bands are presented to each ear (e.g., Gallun, Mason, & Kidd, 2006; Kidd, Mason, & Gallun, 2005).

To ensure that a degradation of the information associated with the stimuli would be detectable, spectrally matched noise was added to all stimuli. Noise levels were chosen independently for each subject and for each task to ensure that baseline performance was sufficiently below perfect.

Listeners were asked either to *identify* keywords in the speech or to simply *detect* the presence of speech in the noise. Listeners participated in three attention conditions. In the first, single-task/single-ear (ST/SE) condition, one stimulus was presented to one ear and the listener performed the same task for a block of 50 trials. In the single-task/dual-ear (ST/DE) condition (selective attention), two stimuli were presented, one to each ear, and the listener performed the same task on the stimuli presented to one of the ears for an entire block of trials. Finally, the dual-task/dual-ear (DT/DE) condition (divided attention) involved the presentation of two stimuli on every trial, with the target ear revealed only after the stimuli had been presented. The task performed on the stimuli presented at a given ear was kept constant throughout a block of trials. Since the only uncertainty that was present (which was present only in the DT/DE condition) was whether the target ear was the right or the left, the listeners could potentially perform the two necessary operations while the stimuli were being presented, encode the appropriate responses, and then make whichever response was cued. This would have resulted in no dual-task costs.

Costs were compared when tasks were the same (both identification, or ID/ID) and when the tasks were different (left-ear detection and right-ear identification; DET/ID). The cost of selective attention was tested by adding a second stimulus at the nontarget ear. This *dual-ear cost* was operationally defined as the difference between listener performance in the ST/SE and ST/DE conditions. The divided-attention cost was tested by asking the listener to perform two tasks within the same trial. This *dual-task cost* was operationally defined as performance in ST/DE minus performance in DT/DE. In order to examine the effects of stimulus length (for reasons described below), costs were measured when the stimuli consisted of full sentences (Experiment 1) as well as when only keywords served as the speech stimuli (Experiment 2).

EXPERIMENT 1

Method

Listeners. Four listeners, 19–32 years of age, participated in 2-h sessions three to four times per week over several months and were paid for their participation. Three of the 4 (L1, L2, and L3) were very familiar with the stimuli because of several thousand trials of listening experience prior to this experiment. All had audiometric thresholds of 20 dB HL or better in each ear for octave frequencies from 250 to 8000 Hz. All subjects reported that they were right-handed.

Stimuli. The speech stimuli were sentences from the coordinate response measure (CRM) corpus (Bolia, Nelson, Ericson, & Simpson, 2000) with the structure "Ready [call sign] go to [color] [number] now," with eight call signs, four colors (white, red, green, and blue), and eight numbers (1–8). The speech was processed to restrict the frequency content, as described in Arbogast et al. (2002). Sentences were passed through a first-order high-pass Butterworth filter with a cutoff of 1200 Hz to roughly equate energy across the spectrum, after which 10 approximately half-octave fourth-order Butterworth filters, evenly spaced on a logarithmic scale from 100 to 3844 Hz, were used to divide the sentence into 10 narrow bands. Half-wave rectification and low-pass filtering at 50 Hz by a fourth-order Butterworth filter extracted the amplitude envelope within each band, reducing the speech waveforms to a set of 10 amplitude envelopes, each associated with 1 of the 10 frequency bands.

Processed speech was generated by randomly choosing 5 of the 10 envelopes and using them to modulate a set of five pure tones with frequencies equal to the center frequencies of the chosen bands. Presenting the five envelope-modulated tones together resulted in a sentence with restricted frequency content and reduced harmonic structure, but with the amplitude variations over time that had occurred in those bands in the original sentence. When presented at an adequate level, such stimuli are perfectly identifiable to listeners (see also Brungart et al., 2005) with only a few minutes of practice—probably due in part to the closed-set nature of the task. Minor variations of this processing scheme have been used in a number of recent informational masking studies (e.g., Arbogast et al., 2002; Brungart et al., 2005; Gallun et al., 2005; Kidd, Mason, & Gallun, 2005) and have been shown to cause very little energetic masking (although in this experiment there is little danger of energetic masking, with stimuli presented to different ears).

For each trial, two sentences were generated in the manner described above, with the requirement that each had a different set of five bands as well as a different call sign (for the complete sentences), color, and number. Each sentence within a pair was also spoken by a different male talker. An example of the frequency content of a pair of companion sentences, one to be presented to the right ear and one to the left ear, appears in Figure 1. It can be seen that the right-ear sentence (gray) and the left-ear sentence (black) each contain mutually exclusive bands and that the peaks are well separated. In an attempt to further simplify the stimulus separation task, the call sign of the right-ear sentence was always constrained to be *Baron*. Sentence lengths were not equalized, but onsets were synchronous.

Masking noise with matched-frequency spectra was generated for all sets of sentences by multiplying each sentence with a random draw of noise in the frequency domain. The noise was thus composed of five narrow bands with center frequencies and bandwidths identical to those of the bands that comprised the sentence from which it had been derived. Simultaneously presenting a processed sentence and its spectrally matched noise reduced the intelligibility (or detectability) of that sentence without affecting the intelligibility of the companion sentence. This allowed baseline performance to be adjusted to fall within an experimenter-defined range.

Psychometric functions and familiarization. As noted, 3 of the listeners had experienced several thousand trials of listening with similar stimuli, but L4 was naive at the beginning of the first experiment. Listeners were familiarized with the stimuli through participation in the ST/SE conditions, starting with stimuli presented in quiet,

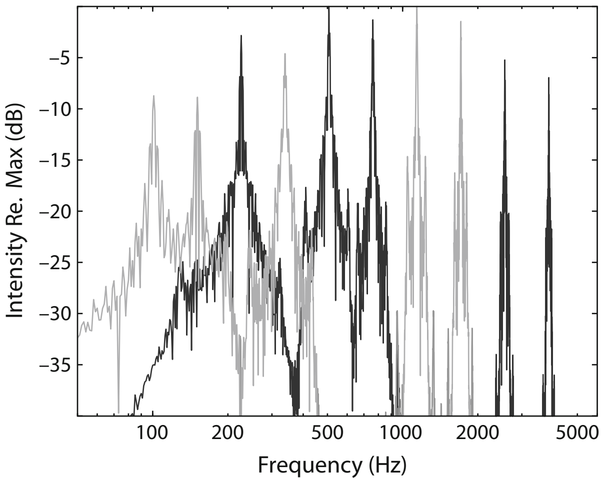

Figure 1. Example of the frequency spectra of one pair of processed-speech stimuli. On each trial, five frequency bands from one sentence were randomly assigned to the left ear (plotted in gray), and the remaining five, from another sentence, were assigned to the right ear (plotted in black). These are the frequency spectra from which the spectrally matched noise would be generated for these two utterances.

and gradually adding noise in order to obtain psychometric functions. The target sentence was always presented at a level of 60 dB SPL, and the noise level was manually adjusted between blocks until performance was in a range of 75%–90% correct. Once the range was established, full psychometric functions were obtained and then fit with a logistic function using the "psignifit" software (bootstrap-software.org/psignifit/), which implements the maximum-likelihood method described by Wichmann and Hill (2001). The integer decibel value that resulted in performance closest to 85% correct was chosen for the ST/SE condition. The psychometric functions were very similar across listeners and quite orderly, allowing a very straightforward determination of the noise levels to be used. For all listeners, one noise level (50 dB SPL) was obtained for the identification task in both the left and right ears. The detection task (performed only on stimuli presented to the left ear) required higher noise levels to reduce performance to criterion. For the experiment, 3 of the listeners were tested at 58 dB SPL and 1 listener (L1) at 60 dB SPL.

Stimulus presentation and data collection. Before each block of 50 trials, an LCD on the response pad indicated whether the target ear was going to be constant throughout the block (and if so, which it would be) or if target ear would vary. The listeners were also told which task(s) they would be asked to perform. Approximately 1 sec after a keypress that initiated the block of trials, the LCD screen became blank and the auditory stimuli were presented. Afterward, a question appeared on the LCD screen that indicated the ear for which a response was required and the task to be performed. For the identification task, the listener made two responses, first indicating the color (four alternatives) and then the number (eight alternatives). For the detection task, listeners made a single response ("yes" or "no"). Responses were entered via a keypad with all of the possible responses written on the keys. Feedback appeared after each trial. Responses were scored as correct or incorrect, with accurate identification of both color and number being required for correct identification scores.

The single- and dual-task conditions were presented in a pseudorandom order, such that no listener finished running one condition more than a few days before finishing the others. Each listener completed a minimum of 200 trials in each condition for each ear, and the ST/SE and ST/DE conditions were repeated throughout the testing to reduce the impact of learning on the data.

Results

Figure 2 shows the results for all 4 listeners from Experiment 1, in which the stimuli were the complete sentences ("Ready [call sign] go to [color] [number] now"). The black bars represent ST/SE performance, the white bars represent ST/DE performance, and the gray bars represent DT/DE performance. Note that the same noise level was used for all three conditions. Panels A and B represent the data for the left-ear (A) and right-ear (B) responses from blocks of trials in which both tasks were identification. Panels C and D represent the data for the left-ear (C) and right-ear (D) responses from blocks in which the task at the left ear was detection and the task at the right ear was identification. Note that the right-ear identification baseline (ST/SE) data are the same for both ID/ID and DET/ID; the data appear in both panels B and D for comparison purposes. Note also that the range on the ordinate is different for the detection and identification tasks.

For ID/ID, the cost of adding a second stimulus to the opposite ear (ST/SE − ST/DE; i.e., the dual-ear cost, or failure of selective attention) was fairly large when the target was at the *left* ear (an average difference in proportion correct of .20), but the cost was smaller when the target was at the *right* ear (a difference in proportion correct of .07). The dual-task cost, or failure of divided attention (ST/DE − DT/DE), was large for both ears (a difference of .23 at the left ear and .24 at the right).

A two-way repeated measures ANOVA on proportions correct for ID/ID examined the main effects of condition and ear of presentation. The effect of condition (ST/SE vs. ST/DE vs. DT/DE) was statistically significant [$F(2,6) = 145.01, p < .001$], as was the effect of the ear to which the target was presented [$F(1,3) = 31.27, p < .05$]. The interaction was not significant, however ($p = .068$). Paired-samples t tests for the dual-ear costs (ST/SE − ST/DE) and dual-task costs (ST/DE − DT/DE) were performed by combining across ears. These analyses showed that both differences were statistically significant ($p < .01$).

For DET/ID, performance in the DT/DE condition was quite comparable to that found in either of the ST conditions. Consequently, the dual-ear cost associated with adding a second stimulus to either ear (ST/SE − ST/DE) was small (an average reduction in proportion correct of .01 for the left ear and .03 for the right). This indicates that listeners were quite successful in selectively attending to only the left or the right ear. For the divided-attention condition, similarly small average dual-task costs were observed (average decrease in proportion correct of .06 for the left ear and .03 for the right), suggesting that the listeners were quite successful in reporting either stimulus as requested. Nonetheless, the individual data (rather than the average costs) suggest that there was a real, if relatively small, dual-task cost in the divided-attention condition for some of the subjects.

A second two-way repeated measures ANOVA, performed on proportions correct in the DET/ID conditions, revealed that the effect of condition (ST/SE vs. ST/DE vs. DT/DE) was statistically significant [$F(2,6) = 25.06$, $p < .001$]. The effect of the ear to which the target was presented was not significant, however [$F(1,3) = 1.25$,

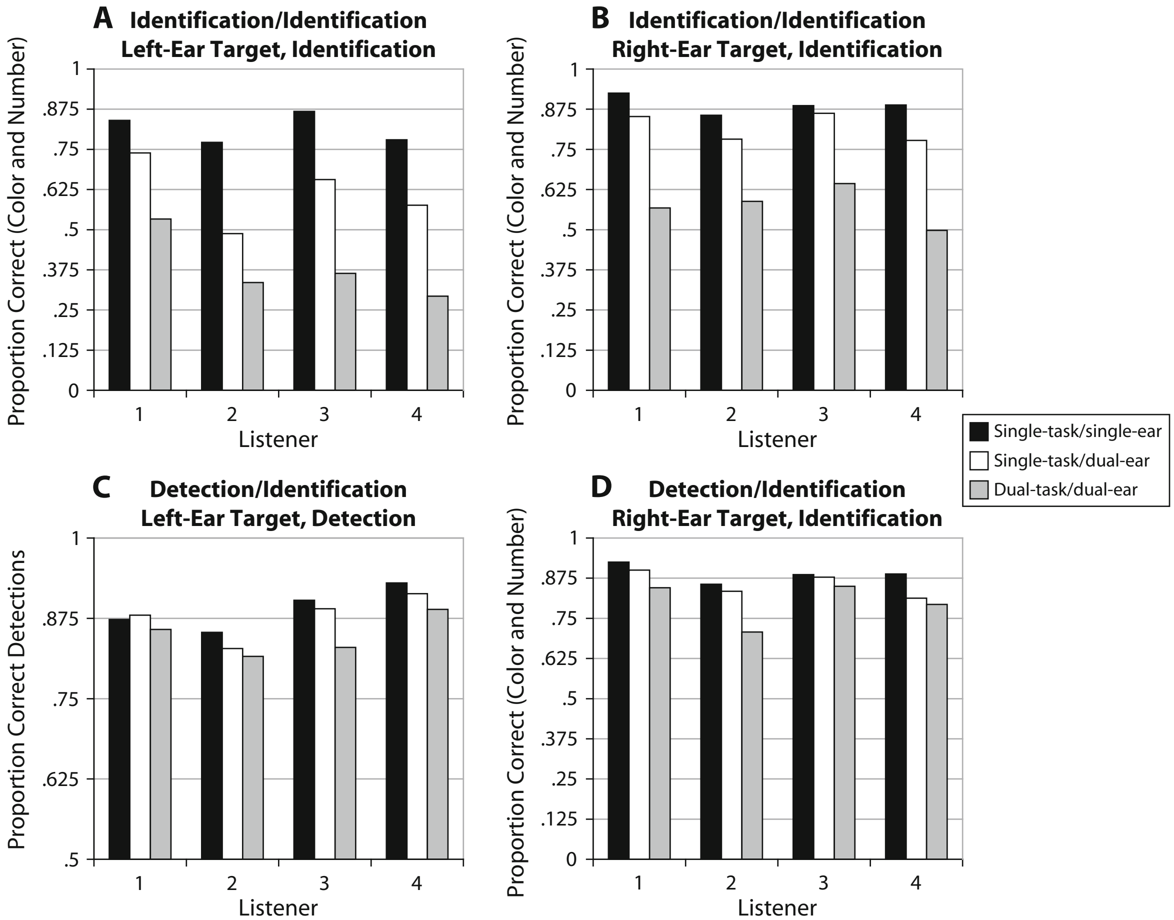

Figure 2. Results of Experiment 1 for all 4 listeners. The bars indicate performance in terms of proportions of correct responses for the single-task/single-ear condition (black bars), the single-task/dual-ear condition (white bars), and the dual-task/dual-ear condition (gray bars). Panels A and B plot the results for the blocks of trials in which the listeners were asked to identify the keywords presented to both the left ear (A) and the right ear (B). Panels C and D plot the results for the blocks in which the listeners were asked to detect the presence of speech in the stimuli presented to the left ear (C) and to identify the keywords presented to the right ear (D).

$p = .346$], nor was the interaction ($p = .283$). Paired-samples t tests for the dual-ear costs (ST/SE − ST/DE) and dual-task costs (ST/DE − DT/DE) were performed by combining across ears. These showed that there was no difference in performance for the two single tasks (ST/SE vs. ST/DE, $p = .079$). On the other hand, when both ears were stimulated, there was a significant difference between the single task and the dual task (ST/DE vs. DT/DE, $p < .05$).

Performance Operating Characteristic (POC)

A graphical representation of the costs is shown in Figure 3, which shows POC graphs (Norman & Bobrow, 1975). The POC (one example of which is the attention operating characteristic of Sperling, 1984) allows performance on two tasks to be compared in an intuitive manner as well as allowing quantitative analysis of three classic divided-attention hypotheses (described below).

The "all-or-none" hypothesis proposes that the observer can process only one input on each trial. In ST/DE, this causes no difficulty, since the target ear is constant for a block of 50 trials. In the DT/DE, however, the observer would have to arbitrarily choose a single stimulus to process on each trial according to this theory, and would lose all information about the unselected stimulus. The prediction is that the listener will choose the incorrect stimulus on half of the trials and will be forced to guess. As long as the probability that either stimulus is the target is equal and the listener truly has no information about which to choose, performance should decline equally in the two tasks. Any bias toward selecting one stimulus over the other, though, would result in a greater decline in performance for one task than for the other.

The quantitative prediction made by the all-or-none hypothesis is that when two tasks must be performed, performance in a DT/DE condition (plotted as a diamond in Figure 3) will be halfway between performance in the ST/DE condition (plotted as squares) and chance performance for each task. This prediction is represented by the dashed line in Figure 3, which will be referred to as the *switch-*

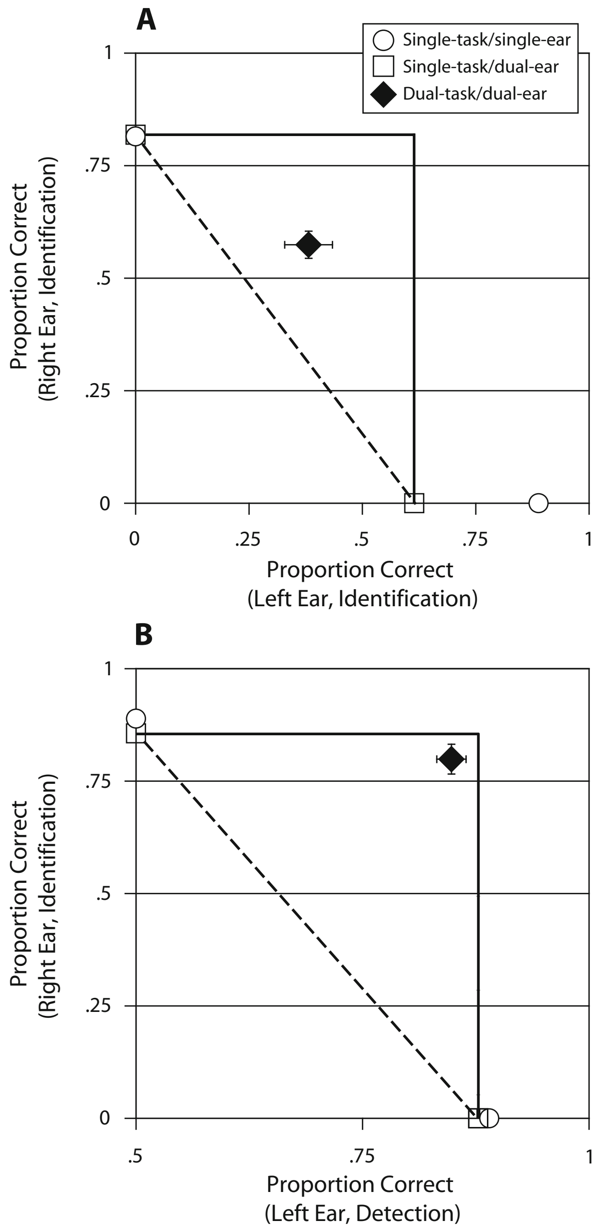

Figure 3. Performance operating characteristics for the averaged results of Experiment 1. Circles represent single-task/single-ear performance, squares represent single-task/dual-ear performance, and diamonds represent dual-task/dual-ear (DT/DE) performance. The dashed line is the predicted DT/DE performance if only one input can be processed on each trial, and the intersection of the solid lines represents complete processing independence (see the text for details). Panel A represents the identification/identification conditions, and panel B represents the detection/identification conditions. Error bars on the DT/DE values reflect standard errors across subjects.

ing line, since it indicates the predicted performance if listeners are forced to switch between the stimuli in an all-or-none fashion. The endpoints of the switching line fall on the ST/DE points, indicating that if one stimulus is chosen on every trial, performance will be as good as it is in ST/DE for that stimulus and at chance on the other. Note that the switching hypothesis takes into account the loss in performance associated with the mere presence of a second stimulus, since it is anchored at the ST/DE points rather than the ST/SE points.

The "independence" hypothesis is that the two tasks involve completely independent resources, so that not knowing the target ear should produce no cost. Consequently, DT/DE performance would fall at the *independence point*, indicated by the point at which the horizontal and vertical solid lines intersect. The independence point, like the switching line, measures dual-task cost after the cost associated with the presence of an interferer has been removed. Because the ST/SE points are also plotted, the interference costs as well as the dual-task cost relative to the single stimulus alone are also displayed.

A third hypothesis is that listeners are not able to process both inputs in parallel, but neither do they lose all information about the unattended stimulus. Two mechanisms that could result in partial information about the unattended stimulus are (1) rapidly switching between stimuli and (2) holding one stimulus in memory while processing the other. These "partial loss" hypotheses are similar to a general quantitative model that has been proposed that likens the division of processing resources to a sampling process in which the observer must time-share between inputs by sampling each for a different proportion of the total time available (Luce & Green, 1978; Norman & Bobrow, 1975). In fact, rapidly switching between inputs could potentially be modeled quite well by a sampling process. The performance of this limited-capacity system should then follow a pattern in which allocating more processing resources (i.e., samples) to one stimulus should improve performance for that stimulus because of a reduction in its associated variance. Unfortunately, testing such a model requires a response measure that is directly related to the variance in the stimulus. Proportion correct is not such a measure, but it can be used to calculate information received, which is the essence of this hypothesis. In the General Discussion below, an analysis of such a divided-resources hypothesis based on information transmission is presented.

Panel A of Figure 3 represents the averaged data from ID/ID, and panel B represents the averaged data from DET/ID. Identification performance in the ST/SE condition at the right ear is based on the same set of data in both panels. Performance from the trials on which the left-ear stimulus was to be reported is plotted on the abscissa, and performance from when the right-ear stimulus was to be reported is plotted on the ordinate. This format makes it easy to see the difference between the ID/ID and DET/ID conditions as well as the degree to which each matches the predictions of independence and of all-or-none switching.

Discussion

The general trend observed in the ID/ID condition is that all 4 listeners performed quite well in the ST/SE condition at either ear (simply showing that appropriate noise levels were chosen), but that performance in the ST/DE condition decreased, especially at the left ear. This finding suggests a dual-ear cost, which indicates a failure of selective attention and is the operational definition of informational masking for this task. That is, listeners had difficulty separating the stimulus at the target ear from that at the nontarget ear, even when the target ear was clearly indicated in advance and did not change throughout a block of trials. This is perhaps a remarkable result, given the separation present in both frequency and ear of presentation. The results of Cherry (1953) and Wood and Cowan (1995), for example, suggest that listeners can always choose to only process the stimuli arriving at one ear. Limiting the stimuli to mutually exclusive frequency regions, as we did, only strengthens the prediction that no dual-ear cost should occur at all. On the other hand, if one considers the requirement that listeners distinguish among 10 different frequency regions and appropriately group 5 into one stimulus and 5 into another, the general success rather than the occasional failures should perhaps be surprising (but not unexpected, given the similar findings of Arbogast et al., 2002; Gallun et al., 2006; and Kidd, Mason, & Gallun, 2005, across a range of numbers of bands). Presumably the division by both ear and frequency leads to a very robust signal in which the 5 bands are easily grouped in the appropriate manner, and the failures that occur are the results of a more central process involving the retrieval and sorting of the words associated with the processed speech.

In the DT/DE condition for the blocks in which listeners were required to identify keywords at both ears (ID/ID), a further drop in performance was observed (a dual-task cost associated with divided attention) for all listeners and both ears. Dual-task costs (indicated by the difference between the DT/DE point and the independence point on the POCs in Figure 3) are the *additional* drop in performance that occurs when the target ear is not indicated in advance and listeners must process the stimuli at both ears in order to achieve optimal performance. The fact that DT/DE performance is not on the switching line argues against an all-or-none allocation of attention to only one stimulus on each trial. Consequently, the results indicate a sharing of resources between the two tasks, but it is not obvious which structural limitation is involved. Although it seems intuitive to point to a speech-specific mechanism, Hafter et al. (1998) attributed similar results with nonspeech stimuli to competition for a limited resource related to the retrieval of context-coded items from long-term memory. Such an explanation is compatible with these data as well, since the mapping between sounds and words clearly depends on information held in long-term memory. If retrieving that information is the limitation on performance for simple intensity identification tasks, there is no reason to believe that there would be less of a limitation when the complexity of the stimulus and of the representation to be retrieved is increased.

In the DET/ID condition, there was very little decrease in performance for conditions intended to produce either selective or divided attention, as indicated by the proximity of both the DT/DE point to the independence point and the ST/DE points to the ST/SE points. This is a remarkable result in light of the difficulties just discussed with dividing attention in the ID/ID condition. These results suggest either that listeners were drawing on two different types of resources that did not compete or that the single resource required was not entirely depleted by these two tasks. Since the identification task and stimuli were identical to those in the ID/ID conditions (with the exception of the noise level, and hence speech intelligibility, at the left ear), it seems most plausible that the two tasks were based on different resources. Perhaps, as Hafter et al. (1998) suggested for their data, the identification task required comparison of the stimulus with a representation stored in long-term memory, whereas the detection task could be performed on the basis of a within-trial comparison. Because adding this type of speech stimulus to noise introduced energy at modulation rates of 2–10 Hz (Kidd, Mason, & Gallun, 2005), the detection task may have been performed as a modulation detection based on the outputs of filters tuned to low modulation rates (Ewert & Dau, 2000; Gallun & Hafter, 2006).

Two additional possibilities could also explain the lack of a dual-task cost for DET/ID. The first is that reducing the intelligibility of the speech at the left ear made these tasks so completely different that even if the listener failed to keep the two ears separated, the tasks could both still be achieved. This argument suggests that the detection task was based on distinguishing one utterance from two and that the identification task suffered no interference because intelligibility of the signal at the left ear was lowered by the use of higher noise levels (the proportions correct were between .4 and .5 when listeners were asked to identify keywords at the noise levels used for detection). The second possibility is that listeners were able to switch their attention between the stimuli during the trial. The optimal switching strategy would involve attending to the left ear for the first portion of the stimulus and then, once a speech detection judgment had been made, switching to the right ear in time to hear the color and number keywords and making an identification judgment. We examined this switching explanation in the second experiment.

For the identification task performed at the right ear, informational masking (which will be operationally defined here as simply the failure of selective attention, ST/SE − ST/DE) was more extensive when paired with an identification stimulus than when paired with a detection stimulus. This result is interesting, because it shows that when the interfering stimulus was presented at a *lower* level (identification stimulus, noise at 50 dB, speech at 60 dB SPL), it was more effective than when a noisier stimulus was presented at a higher level (detection stimulus, noise at 58 or 60 dB, speech at 60 dB SPL). Although the difference between the identification stimulus and the combined speech-plus-noise stimulus was only a few decibels, this still suggests that the interference was truly informational (rather than caused by energetic overlap at a peripheral analyzer), since the softer (but more intelligible) signal caused more interference. A similar effect of reducing the effectiveness of interfering speech by adding noise was described by Kidd, Mason, and Gallun (2005).

The amount of informational masking at the left ear (again, operationally defined as ST/SE − ST/DE) was much greater for the identification task than for the detection task, despite the presence of identical stimuli at the right ear in both cases. This supports the suggestion that the drop in performance was due to informational rather than energetic masking (since the amount of energetic masking was identical for the two tasks) and indicates that a single type of masking stimulus can exert different influences on different targets. The large increase in errors for the identification task at the left ear relative to all of the other selective-attention conditions is not surprising, given the reports in the literature of the dominance of the right ear for speech processing (e.g., Kimura, 1961; Milner, Taylor, & Sperry, 1968; Studdert-Kennedy & Shankweiler, 1970). It is interesting that there was essentially no interference for the speech *detection* task, however. This suggests that perhaps listeners were indeed using different processing mechanisms when they were performing the identification and the detection tasks.

Recent neurophysiological work has shown support for the hypothesis, based on both psychophysical results and lesion studies, that speech is preferentially processed by the left hemisphere (e.g., Giraud et al., 2000; Narain et al., 2003; Scott & Johnsrude, 2003; Tervaniemi & Hugdahl, 2003.) A paired-samples t test on the dual-ear costs at the left and right ears for ID/ID found that the difference was significant ($p = .023$), but a similar analysis on the dual-task costs for ID/ID found that the difference was not significant ($p = .45$). These effects can be seen most clearly in the POC in Figure 3A, where the DT/DE point is not shifted toward the ordinate or the abscissa, but the difference between the ST/DE and DT/DE points is greater on the abscissa (left-ear task) than on the ordinate (right-ear task). That performance was reliably different across ears in the selective- but not the divided-attention task suggests that the deficit in speech-processing ability does not create an additional burden in the case of divided attention once the cost in the selective condition has been accounted for. This is similar to suggesting that the speech presented to the left ear is slightly degraded (or more subject to intrusions from the right-ear stimulus) but that there is no additional cost associated with dividing attention to degraded and nondegraded speech. These results are also consistent with the finding in the literature that differences between performance at the left and right ears are most obvious in conditions of selective attention (e.g., Milner et al., 1968).

EXPERIMENT 2

The large dual-task costs found in Experiment 1 for the ID/ID conditions resemble the interference reported by authors such as Treisman (1964) and Hirst and Kalmar (1987) for stimuli composed of spoken sentences and by Bonnel and Hafter (1998) and Hafter et al. (1998) for the simultaneous identification of the direction of changes in intensity of a tone and a light. Such interference was not found by authors such as Shiffrin et al. (1974) for stimuli composed of only a few syllables. This effect of stimulus length or complexity raises the possibility that if the stimuli were shortened to contain only the keywords, listeners might experience smaller dual-task costs when the tasks performed at the two ears were the same. One reason performance is poorer in the dual-task condition for longer stimuli may be that the memory load is higher. Although the response required probably did not reach the limits of the memory system (Cowan, 1984, 1988, 1995, 2001; Parkinson, 1974; Parkinson, Knight, DeMaio, & Connors, 1974), it is possible that a memory limitation is associated with presenting relevant stimuli among irrelevant stimuli. Essentially, if the memory system relies on a "buffer" that is automatically filled with whatever stimuli are presented (Cowan, 1984, 1988, 1995, 2001), even if the relevant keywords do not fill the buffer, the rest of the sentence might do so. For this reason, stimuli shortened to only the keywords were used in Experiment 2.

Another reason to alter the stimuli involves the DET/ID dual-task condition. As mentioned above, it is theoretically possible that the small dual-task costs found when the tasks at the two ears were different were a result of a *within-trial* switching strategy. By shortening the stimuli to just the keywords, such a strategy would become more difficult to implement. As Miller and Bonnel (1994) noted, however, at a sufficiently rapid rate it becomes impossible to distinguish between a switching strategy and truly independent processing.

Method

Three listeners from the first experiment also participated in the second (L2, L3, and L4). For the new experiment, the processed sentences were reduced in length by editing the digital waveforms of the original sentences to remove the portion of the waveform that corresponded to the words "Ready [call sign] go to" and "now," leaving just the color and number keywords. The psychometric functions (not shown) measured to determine the appropriate noise levels were much more variable for Experiment 2, both within and across listeners (despite similar numbers of trials and the fact that all subjects had just completed Experiment 1). Unlike in Experiment 1, one of the listeners (L4) was not even able to produce a left-ear identification function in which performance reliably decreased as noise level was increased.[1] The detection functions were less variable than the identification functions, however. On the basis of the psychometric functions, the spectrally matched noise in the identification task was presented at 40 rather than 50 dB SPL. In the detection task, the noise level was set at 55 dB SPL. In all other respects, the stimuli and methods of Experiment 2 resembled those of the first experiment.

Results

Figure 4 represents the results of Experiment 2 in the same manner that Figure 2 represented those of Experiment 1. The black bars indicate ST/SE performance, the white bars indicate ST/DE performance, and the gray bars indicate DT/DE performance. Overall, performance was worse for the 3 listeners in Experiment 2, despite attempts to equalize performance by using less intense noise than in Experiment 1.

The dual-ear costs (ST/SE − ST/DE) were much larger when the identification task was being performed on the stimuli at the left ear (average drop in proportion correct of .20) than when the task was to attend to the right ear (drop in proportion correct of .08). For the divided-attention task, the cost (ST/DE − DT/DE) was again greater for the left ear (with a drop of .16) than for the right (drop of .12).

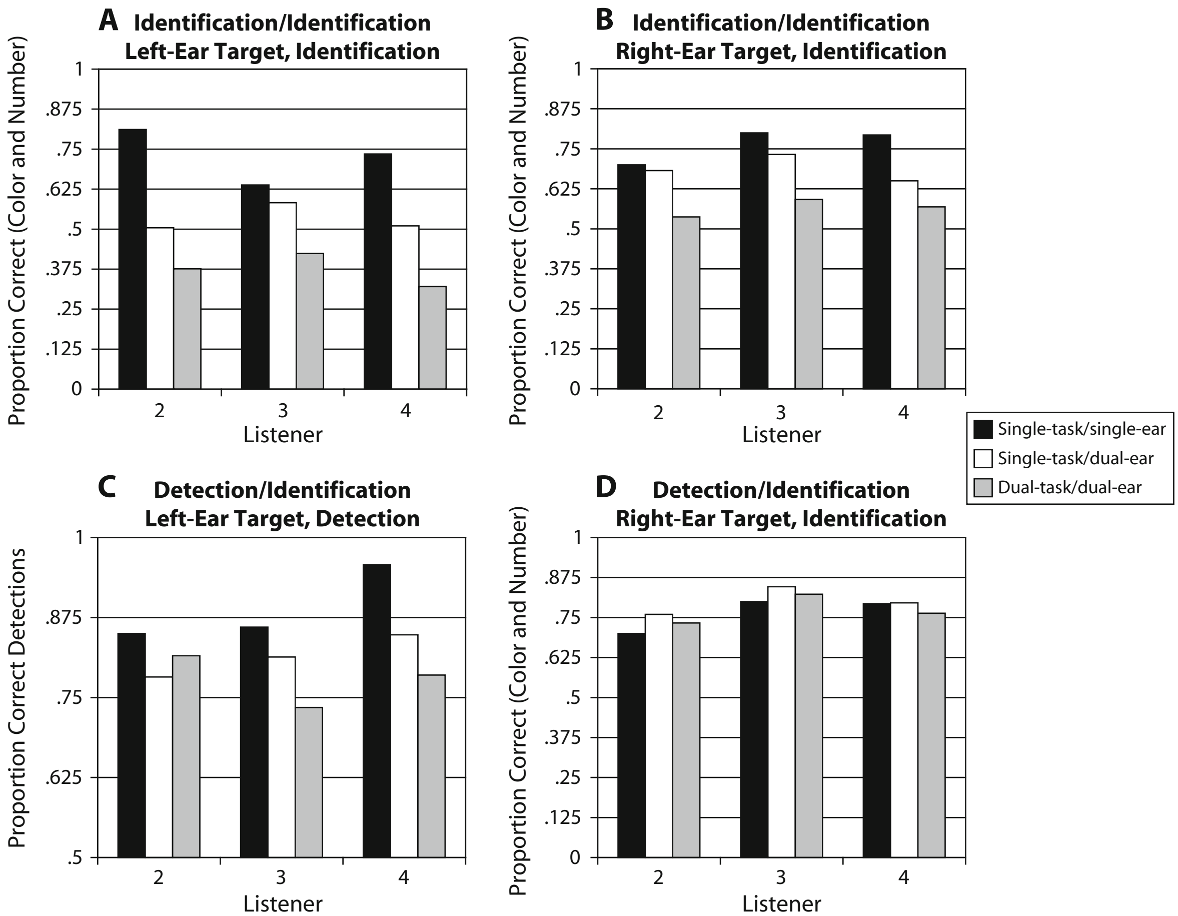

Figure 4. Results of Experiment 2 for all 3 listeners. The organization and coding is the same as in Figure 2.

A two-way repeated measures ANOVA was performed on the proportion correct data plotted in panels A and B of Figure 4 (ID/ID). The main effect of condition (ST/SE vs. ST/DE vs. DT/DE) was statistically significant [$F(2,4) = 44.88$, $p = .002$], as was the effect of the ear to which the target was presented [$F(1,2) = 23.82$, $p = .04$]. The interaction, though, was not significant ($p = .192$). Since there was no interaction, paired-samples t tests for the dual-ear costs (ST/SE − ST/DE) and dual-task costs (ST/DE − DT/DE) were performed by combining across ears. There was no significant dual-ear cost ($p = .069$), but the dual-task cost was significant ($p = .001$).

Although substantial, the dual-ear cost was not significant because of differences across listeners. As can be seen in panels A and B of Figure 4, L2 has a highly asymmetric dual-ear cost (with costs of .30 for the left and .02 for the right ear), L3's dual-ear costs are symmetrically distributed and small (approximately .06 for both ears), and L4's dual-ear costs are substantial for both ears (.22 on the left and .14 on the right).

These same variations across listeners also impacted the reliability of the effects of ear of presentation that were observed in Experiment 1. Although the main effect of ear was significant ($p = .04$), a paired-samples t test on the dual-ear cost at the left and right ears revealed that the difference between ears was not significant ($p = .288$), nor was the difference in dual-task costs for the two ears ($p = .727$). As with the dual-ear costs detailed above, the lack of statistical reliability is a direct consequence of the fact that ear-of-presentation resulted in costs for some listeners but not for others. This heterogeneity of results is typical of studies of informational masking, and it represents an important theoretical problem associated with fitting a computational model to the results (see, e.g., Durlach et al., 2005; Lutfi, 1993).

Panels C (left ear) and D (right ear) of Figure 4 demonstrate that adding a stimulus to the other ear in the DET/ID conditions reduced performance in left-ear detection (ST/SE − ST/DE; average drop in proportion correct of .07) much more than in right-ear identification (*increase* in proportion correct of .04). The divided-attention cost (ST/DE − DT/DE) was slight for both the left-ear detection task (drop of .04) and the right-ear identification task (drop of .03). This suggests that shortening the stimuli did *not* increase the cost associated with dividing attention in order to perform the DET/ID dual-task condition, especially for the right-ear identification task. This result is not consistent with the prediction of the switching hypothesis.

A two-way repeated measures ANOVA performed on the DET/ID data, obtained when the tasks at the two ears were different, found no significant main effect of condition (ST/SE vs. ST/DE vs. DT/DE) [$F(2,4) = 2.77$, $p = .18$], nor an effect of the ear at which the target was presented [$F(1,2) = 1.96$, $p = .30$]. There was, however, a significant interaction [$F(2,4) = 13.22$, $p = .017$]. Paired-samples t tests revealed the source of the interaction to be the ST/SE detection condition at the left ear, which was reliably different from detection performance in the ST/DE and DT/DE conditions at a level of $p < .05$. There was no significant difference between the ST/DE and DT/DE conditions for the detection task performed at the left ear, however. Nor were the differences significant for any of the conditions when the task was identification at the right ear. Because of the interaction based on ear of presentation, no analyses were performed on the data combined across ears. These results confirm the trends apparent in panels C and D of Figure 4: The only significant differences in the DET/ID conditions for Experiment 2 were due to the dual-ear costs for detection at the left ear.

Discussion

The results of Experiment 2 confirm those of Experiment 1 and, by using shorter stimuli, extend the earlier findings by providing evidence against the hypothesis that listeners obtained nearly independent performance in the DET/ID task by switching attention between ears midway through the stimulus. A fast-switching strategy (operating perhaps at the 6-Hz rate proposed by Broadbent, 1958) still cannot be ruled out, however. That the shorter stimuli produced the same pattern of results also suggests that listeners are not limited in the ID/ID dual-task condition by interference from the irrelevant words that come before and after the keywords. In addition, these results address any concerns related to the asynchrony between keywords introduced by differences in the rates at which the sentences were spoken. When only the keywords were used, it was necessary for the listeners to process both inputs essentially simultaneously.

GENERAL DISCUSSION

Variance-Based Limitations on Performance

Listeners in these experiments performed as if they either were sharing a single resource (ID/ID) or had access to independent resources (DET/ID). As mentioned above, the division of processing resources can be modeled as a sampling process, in which the total number of samples (and thus the information) is fixed (Luce & Green, 1978; Norman & Bobrow, 1975). Using this idea, Bonnel and Miller (1994) fit a variance-based model of performance to a set of results from a dual-task condition in which listeners were asked to allocate various proportions of their attention to one stimulus or the other. By introducing several levels of stimulus variance, Bonnel and Miller were able to demonstrate that allocation of attention could be modeled as reducing a source of *internal* variance that was proportional to the variance in the stimulus. Whereas they speculated that this source might be criterion variance, other researchers (Lu & Dosher, 1998, 1999) have suggested that in addition to additive internal noise, there is a multiplicative internal noise that scales with external noise. Both of these sources of internal variance have been shown to be reducible by attention (Dosher & Lu, 2000; Lu & Dosher, 1998). Bonnel and Hafter (1998) and Hafter et al. (1998) applied the sample-size model to their data and found that when asked to identify a visual as well as an auditory change in intensity, observer performance was well fit by a model that suggested that a limited number of samples were allocated to each stimulus. When asked simply to detect an intensity change in both modalities, listeners behaved as if they were not losing any samples by performing the two tasks simultaneously.

Fitting a variance-based model depends on the use of a performance measure that scales linearly with changes in variance. Previous work, such as that of Bonnel and her colleagues (Bonnel & Hafter, 1998; Hafter et al., 1998), has relied on measurements of d' in two-alternative procedures. Although such a measurement could be applied to the detection data in this study, the identification data (from a 32-alternative task) are not as amenable to such an analysis. One might use the 32-alternative forced choice approximation given by Green and Dai (1991), but it is not clear that such an approximation is appropriate for a situation in which the confusions between the various alternatives are based on vocal parameters that do not lend themselves to completely random errors. In addition, the fact that there are four colors and eight numbers to distinguish means that difficulties with the color words will have a greater influence on performance than will difficulties with the number words.

An alternative would be to use the metric of information received by the listener, measured in bits. As originally laid out by Lindsay and his colleagues (Lindsay, Cuddy, & Tulving, 1965; Lindsay, Taylor, & Forbes, 1968; Tulving & Lindsay, 1967), the shared-resources hypothesis postulates that the limitation on dividing attention can be modeled as an upper limit on the total number of samples that can be extracted by a human observer. It was this hypothesis that led to the predictions of Bonnel and Miller (1994), Bonnel and Hafter (1998), and Hafter et al. (1998). For these experiments, the prediction of this hypothesis would be that, when the same resource is being used, the total proportion of information obtained from each channel should be reduced by a fraction that corresponds to the amount of resources allocated to the competing channel. Assuming that listeners divided their attention equally (which they were instructed to do), the total amount of information received should drop by 50%. When different resources are being used, however, the proportion of total information received should remain essentially unchanged, whether or not the listener knows in advance which channel to report.

The data from Experiments 1 and 2 provide strong support for the prediction that using different resources leads to little or no drop in information. In the selective-attention condition (ST/DE) for the DET/ID conditions, the total information received was 5.59 bits for the long sentences and 5.38 for the short (perfect performance would have been 6 bits: 5 from identification and 1 from detection).

In the divided-attention condition (DT/DE), performance was nearly the same, with an average of 5.43 bits received for the long sentences versus 5.26 for the short ($p < .03$ and $p = .2$, respectively, for paired-samples t tests, ST/DE vs. DT/DE). In terms of the proportion of information received relative to that received in the selective condition, 97% was received for the long sentences and 98% for the short.

On the other hand, the information analysis provides only weak support for the prediction that sharing resources can be modeled as dividing a fixed number of samples. In the selective-attention condition (ST/DE) for the ID/ID conditions, the total number of bits received in the left and right channels combined was 8.99 for the long sentences and 8.55 for the short sentences (perfect performance would have been 10 bits). If dividing attention reduces the total information received by 50%, information received in the divided-attention condition (DT/DE) should have dropped to 4.49 bits and 4.27 bits, respectively. This prediction was not supported by the data, since the total bits received dropped, but only to 7.77 and 7.75, respectively ($p < .001$ for both in paired-samples t tests, ST/DE vs. DT/DE). This does not support the quantitative prediction that only one stimulus could be processed at a time and that any unprocessed information was lost; rather than receiving only 50% of the information presented in the selective condition, listeners received on average 86% and 91% for the two types of stimuli.

The analysis of the dual-task costs in terms of information suggests that listeners in the ID/ID conditions were able to receive a substantial amount of information from both channels. Thus, although there is evidence of a shared resource, the sharing does not seem to act in an all-or-none fashion. This is an important point, and one that might not be as obvious if the data had only been analyzed in terms of proportions correct. Alternatively, if the question is whether the same levels of performance can be maintained when two tasks are required, the proportion correct metric is probably more informative, since the loss of a single bit drops proportions correct from 1.00 to .50.

Alternative Models

An alternative explanation for the dual-task costs in the ID/ID conditions is the hypothesis that listeners were simply not obliged to process the entire utterance in the given time period, since they could have relied on short-term memory. Cowan (1984, 1988, 1995, 2001) has suggested that there exists a sensory buffer that can retain information for several seconds (perhaps up to 10), provided that no new information is presented. Cowan distinguishes between information that is being actively processed and information that is being held in the sensory buffer for later processing. This is in agreement with the findings of Wood and Cowan (1995) as well as with the (visual) memory consolidation theory of Vogel and Luck (2002). Consequently, listeners may have stored one input in an unprocessed, unlabeled "sensory" form while processing the other to the point of labeling with a more easily remembered word (such as *blue* or *four*). This would have allowed them to then store the labeled output and begin processing the contents of the storage buffer once the shared resource was available. Seen from that perspective, it is surprising that our participants experienced any loss of information at all. One might even suppose that what loss was observed was due entirely to failures of memory rather than to demand for a shared speech-processing resource.

A slightly different version of the sensory memory explanation includes the hypothesis that information gradually degrades while held in the sensory buffer and/or may be replaced by subsequently presented auditory stimuli (Cowan, 1984, 1988, 1995, 2001; Vogel & Luck, 2002). If this hypothesis is correct, then it is not too surprising that performance declined in the dual-task condition. In order to examine this hypothesis in detail, it would be useful to adopt the poststimulus masking paradigm often employed in vision studies, in order to ensure that observers are not basing responses on sensory persistence (Sperling, 1960). Since the time course of audition is more rapid than the time course of vision, such measures have not generally been employed. Results such as those on backward recognition masking (Massaro, 1975), however, support the idea that sensory memory may routinely play a role in psychoacoustical tasks, whether this role is acknowledged or not. For further discussion of these issues, see Cowan (1984, 1988, 1995, 2001).

One alternative to a memory explanation is based on the "perceptual load" theory of Lavie and colleagues (for a review, see Lavie, 2005). Lavie and Tsal (1994) proposed that interference found in selective-attention tasks is based on the existence of a *lower* limit on resource allocation. That is, the perceptual system automatically allocates resources to processing stimuli, regardless of whether a higher-level prioritization scheme has designated those stimuli as targets or distractors. Thus, if a task is not sufficiently resource demanding, additional resources will be automatically allocated to irrelevant stimuli. If those irrelevant stimuli compete with the target at a later stage of processing, interference will then occur. Thus, interference will be greatest when the target and masker are clearly separable (so that substantial processing resources are still available) and when the response appropriate to the masker is one of the possibilities for the target. Lavie, Hirst, de Fockert, and Viding (2004) expanded this hypothesis to include cognitive control issues as well.

Applying Lavie's (2005) hypothesis to these results leads to the suggestion that the interference observed in the ID/ID selective-attention conditions was due to automatic processing of the distractor sentence. Since color and number words from both ears were available, interference occurred at the point of response selection. This is actually a fairly compelling explanation from a phenomenological perspective, since listeners often report the experience of hearing only a single set of keywords, which are sometimes those associated with the distractor. This explanation is also in agreement with the quantitative results of Brungart et al. (2001), who found that over 90% of the errors in their study involved erroneously reporting the nontarget keywords. For our study, however, the rate was only 65%. This difference could be a result of the addition of noise to the stimuli, which decreased intelligibility and may thus have resulted in a larger proportion of random errors. Error patterns are discussed in more detail below.

Comparisons With Other Work

In the ID/ID dual-task conditions, listeners were unable to fully process both streams simultaneously. Nonetheless, they were able to report some information about both streams. This contradicts some of the findings of a recent study with *three* different speech stimuli (Kidd, Arbogast, et al., 2005). In one condition of that study, listeners were presented with three simultaneous utterances and no pretrial cues as to which was the target. In that case, listeners behaved as if they were extracting information from only one speech source, since performance was only one-third as good as when the target was identified in advance. The researchers tested a model based on the hypothesis that listeners could only process a single utterance at any one time but were able to switch from one utterance to another within a trial. They found that the proportion correct results matched the predictions of the model, but the error patterns did not. In particular, there were situations in which the listeners were predicted to switch to another speech stream, but the errors showed that on approximately half of the trials the listeners had not. In order to understand the differences between our study and Kidd, Arbogast, et al. (2005), it may be important to consider how the number of talkers and their separability by spatial position influences and limits the strategies available to listeners. It may be a very different situation to be presented with three natural utterances played simultaneously in a sound field than to hear only two utterances, both of which have been processed into mutually exclusive bands, with only one utterance and set of bands present at each ear.

Another aspect of these data that can be compared with what has been found in other studies is the pattern of errors. This aspect is relevant to the issues being considered here, since one logical explanation for the dual-task costs in this study involves failures not of selecting the proper stimulus, but of remembering the correct response. The distinction concerns whether listeners can select a single input for processing to the point of recognition (Broadbent, 1958; Treisman, 1969), or whether both stimuli are processed to the point of recognition (or "labeling") and then the selection involves which stimuli are reported. Unfortunately, distinguishing between these two possibilities is quite difficult, because either a failure of selection for processing or a failure of selecting a response would lead to responses that matched keywords presented to the nontarget ear. Several studies using similar sentence materials have explicitly looked at the source of errors in a multiple-talker situation (Arbogast et al., 2002; Brungart & Simpson, 2002; Brungart et al., 2001; Kidd, Arbogast, et al., 2005) and have concluded that the great majority (often over 90%) of the errors come from nontarget (masker) keywords presented on that trial. As mentioned above, however, an analysis of the errors in the present experiments showed that, although the percentages of errors that included either of the keywords spoken at the nontarget ear were both remarkably similar between the two experiments and much higher than would occur by chance (3.5%), their values were uniformly below 65% for all listeners in all conditions. The fact that these percentages were so much lower than is usually found in such studies was probably due to the spectrally matched noise that had been added to the stimuli, although it is not clear why reducing intelligibility by 15% would reduce intrusions by 35% (and even more for some listeners in some conditions; see below). This suggests that response competition may not account for all of the difficulties encountered by our listeners. The alternative is that listeners actually selected the wrong sentence to process. If they were occasionally aware of this error, and thus purposefully guessed a different color and number, the proportion of errors matching the nontarget sentences would be reduced. This explanation suggests that there is a difference between actively selecting the wrong sentence and simply suffering confusion over which sentence to process and/or to report.

In the ID/ID conditions, on DT/DE trials 53% of the errors on the left-ear identification task (averaged across all listeners and both experiments) involved incorrectly reporting a right-ear color or number, but only 31% of the errors on the right-ear identification task involved a word that had been spoken at the left ear. Likewise, on ST/DE trials, over 52% of the errors for left-ear targets were words that had been presented to the right ear, but only 4% of the errors for right-ear targets were words that had been spoken at the left ear. It is possible that speech presented to the right ear gains preferential access to the speech identification mechanism, and thus might be more easily processed than speech presented to the left ear (see, e.g., Kimura, 1961; Milner et al., 1968; Studdert-Kennedy & Shankweiler, 1970). The differences we found between errors due to intrusions from the right ear and from the left ear support this interpretation.

A similar error analysis of the DET/ID conditions found that the percentage of trials on which the responses in the identification task matched both keywords from the nontarget ear was never above 7%—a result similar to what would have occurred by chance (5%).

In terms of relating these results to the literature on divided attention, the work of Bonnel and Hafter (1998) and Hafter et al. (1998) is particularly relevant. In those studies, interference was associated with identification tasks and noninterference with detection. Although, unlike theirs, our study involved speech stimuli, it also involved identification and detection. In fact, it is possible that both the detection of intensity increments and the detection of speech rely on the same mechanism: the detection of low rates of amplitude modulation. Kidd, Mason, and Gallun (2005) showed that sentences processed in a manner very similar to those used in this study differ from spectrally matched noise in terms of the relative energy present at low modulation rates (2–10 Hz). Consequently, the detection task could have been performed by detecting a feature of the envelope spectrum that was correlated with the presence of speech. If listeners relied on the output of a modulation filter tuned to 4 Hz, for example, there might have been no need to access long-term memory, and thus (perhaps) no interference with the identification task. Alternatively, modulation detection may have involved accessing long-term memory in a manner that could be performed in parallel with speech identification. Intensity detection of the sort studied by Hafter and his colleagues has also been shown to be compatible with detecting the presence of modulation in the region around 4 Hz (Gallun & Hafter, 2006). Future work might benefit from considering the possible influence of sensitivity to modulation whenever amplitude-modulated stimuli are employed.

The conjecture of Hafter et al. (1998) that dual-task costs are the result of interference associated with particular processing mechanisms is a general insight that may turn out to be applicable to a wide variety of tasks. In their work, the mechanism was associated with the identification of a change in intensity, whereas in the study reported here it was associated with the identification of speech. In both cases, it is reasonable to conclude that a call to long-term memory was involved. The tasks that did not result in interference may both have been performed on the basis of sensitivity to amplitude modulation (as has been suggested for the detection of intensity changes in the work of Gallun & Hafter, 2006). Whether the interference in both studies is tied to the same source (reliance on long-term memory) or to different sources (intensity identification vs. speech identification), the significant finding is that the source of interference in the dual-task conditions lies more in the tasks than in the stimuli. In addition, it is important to note that our experiments did not find support for a single processing-resource bottleneck through which only one input could flow at a time. Although information transmission was decreased by a significant amount, it was still high enough that listeners seem to have been processing both inputs "at the same time," just with reduced efficiency when the same processing resources were required. We suggested that reliance on a sensory memory buffer could have facilitated this ability.

A specific hypothesis that is compatible with these explanations is that the mechanisms responsible for speech processing and amplitude modulation processing may be anatomically distinct. Functional magnetic resonance imaging work over the past decade has produced a number of results that confirm the existence of brain areas that carry out speech-specific processing (reviewed in Narain et al., 2003), as well as suggesting that there may be independent areas of the brain that are sensitive to amplitude modulation (Giraud et al., 2000; Scott & Johnsrude, 2003). From this perspective, it seems plausible that the identification and detection tasks used in this study relied on separate and independent mechanisms. Consequently, when two tasks call on the same mechanism, performance suffers, whereas when they call on different mechanisms (and/or fail to deplete shared resources), little cost is involved for performing the two tasks simultaneously.

Summary

In two experiments, we presented speech that had been processed into two sets of narrow, mutually exclusive frequency bands to opposite ears of participants. When listeners were asked to identify keywords contained in a target sentence, performance was better when the target ear was known in advance than when it was known only after both stimuli had been presented. This performance difference was largely eliminated when listeners were asked to identify keywords in a target sentence presented to the right ear and to detect the presence of speech in the stimulus presented to the left ear. Spectrally matched noise was added to all stimuli at a level that roughly equated performance across ears, listeners, and tasks when a single stimulus was presented. These results show that (1) distinguishing the costs of selective attention and divided attention is informative, and (2) under conditions of divided attention, with very similar stimuli, it is possible to find substantial costs when the tasks are similar or essentially no cost when the tasks are dissimilar. Further work is needed in order to expand the types of tasks and stimuli for which the costs of selective and divided attention have been measured in the same experiment, as well as to test the validity and generality of the specific hypotheses generated to explain our present data.

AUTHOR NOTE

This work was supported by Grants DC00100, DC04545, and DC04663 from the NIH/NIDCD and by AFOSR award FA9550-05-1-2005. F.J.G. was supported by Grant F32 DC006526 from NIDCD. Portions of this research were presented at the 2005 Midwinter Meeting of the Association for Research in Otolaryngology and the 149th Meeting of the Acoustical Society of America. The authors are grateful to Barbara Shinn-Cunningham for comments on an earlier version of the manuscript, to Kelly Egan and Jackie Therieau for their assistance in data collection, and of course to our listeners. Correspondence relating to this article may be sent to F. J. Gallun, National Center for Rehabilitative Auditory Research, Portland VA Medical Center, 3710 SW US Veterans Hospital Road (NCRAR), Portland, OR 97239 (e-mail: frederick.gallun@va.gov).

REFERENCES

Arbogast, T. L., Mason, C. R., & Kidd, G., Jr. (2002). The effect of spatial separation on informational and energetic masking of speech. *Journal of the Acoustical Society of America*, **112**, 2086-2098.

Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *Journal of the Acoustical Society of America*, **107**, 1065-1066.

Bonnel, A.-M., & Hafter, E. R. (1998). Divided attention between simultaneous auditory and visual signals. *Perception & Psychophysics*, **60**, 179-190.

Bonnel, A.-M., & Miller, J. (1994). Attentional effects on concurrent psychophysical discriminations: Investigations of a sample-size model. *Perception & Psychophysics*, **55**, 162-179.

Broadbent, D. E. (1958). *Perception and communication*. New York: Pergamon.

Brungart, D. S., & Simpson, B. D. (2002). Within-ear and across-ear interference in a cocktail-party listening task. *Journal of the Acoustical Society of America*, **112**, 2985-2995.

Brungart, D. S., Simpson, B. D., Darwin, C. J., Arbogast, T. L., & Kidd, G., Jr. (2005). Across-ear interference from parametrically-degraded synthetic speech signals in a dichotic cocktail-party listening task. *Journal of the Acoustical Society of America*, **118**, 292-304.

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *Journal of the Acoustical Society of America*, **110**, 2527-2538.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America*, **25**, 975-979.

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin*, **96**, 341-370.

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

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. *Behavior & Brain Sciences*, **24**, 87-185.

Dosher, B. A., & Lu, Z.-L. (2000). Noise exclusion in spatial attention. *Psychological Science*, **11**, 139-146.

Durlach, N. I., & Braida, L. D. (1969). Intensity perception: I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, **46**, 372-383.

Durlach, N. I., Mason, C. R., Gallun, F. J., Shinn-Cunningham, B., Colburn, H. S., & Kidd, G., Jr. (2005). Informational masking for simultaneous nonspeech stimuli: Psychometric functions for fixed and randomly mixed maskers. *Journal of the Acoustical Society of America*, **118**, 2482-2497.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., & Kidd, G., Jr. (2003). Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target–masker similarity. *Journal of the Acoustical Society of America*, **114**, 368-379.

Ewert, S. D., & Dau, T. (2000). Characterizing frequency selectivity for envelope fluctuations. *Journal of the Acoustical Society of America*, **108**, 1181-1196.

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *Journal of the Acoustical Society of America*, **109**, 2112-2122.

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *Journal of the Acoustical Society of America*, **115**, 2246-2256.

Freyman, R. L., Helfer, K. S., McCall, D. D., & Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *Journal of the Acoustical Society of America*, **106**, 3578-3588.

Gallun, F. J., & Hafter, E. R. (2006). Amplitude modulation as a mechanism for increment detection. *Journal of the Acoustical Society of America*, **119**, 3919-3930.

Gallun, F. J., Mason, C. R., & Kidd, G., Jr. (2005). Binaural release from informational masking in a speech identification task. *Journal of the Acoustical Society of America*, **118**, 1614-1625.

Gallun, F. J., Mason, C. R., & Kidd, G., Jr. (2006). Contralateral masking of narrow bands of speech. *Association for Research in Otolaryngology Abstracts*, 558. Available at www.aro.org/archives/2006/2006_558.html.

Giraud, A.-L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., & Kleinschmidt, A. (2000). Representation of the temporal envelope of sounds in the human brain. *Journal of Neurophysiology*, **84**, 1588-1598.

Green, D. M., & Dai, H. (1991). Probability of being correct with 1 of *M* orthogonal signals. *Perception & Psychophysics*, **49**, 100-101.

Hafter, E. R., Bonnel, A.-M., Gallun, E., & Cohen, E. (1998). A role for memory in divided attention between two independent stimuli. In A. R. Palmer, A. Rees, A. Q. Summerfield, & R. Meddis (Eds.), *Psychophysical and physiological advances in hearing* (pp. 228-238). London: Whurr.

Hirst, W., & Kalmar, D. (1987). Characterizing attentional resources. *Journal of Experimental Psychology: General*, **116**, 68-81.

Kidd, G., Jr., Arbogast, T. L., Mason, C. R., & Gallun, F. J. (2005). The advantage of knowing where to listen. *Journal of the Acoustical Society of America*, **118**, 3804-3815.

Kidd, G., Jr., Mason, C. R., & Gallun, F. J. (2005). Combining energetic and informational masking for speech identification. *Journal of the Acoustical Society of America*, **118**, 982-992.

Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (in press). Informational masking. In W. Yost (Ed.), *Auditory perception of sound sources* (Springer Handbook of Auditory Research, Vol. 29). New York: Springer.

Kimura, D. (1961). Cerebral dominance and the perception of verbal stimuli. *Canadian Journal of Psychology*, **15**, 166-171.

Lavie, N. (2005). Distracted and confused? Selective attention under load. *Trends in Cognitive Sciences*, **9**, 75-82.

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, **133**, 339-354.

Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & Psychophysics*, **56**, 183-197.

Lindsay, P. H., Cuddy, L. L., & Tulving, E. (1965). Absolute judgments of simultaneously presented visual and auditory stimuli. *Psychonomic Science*, **2**, 211-212.

Lindsay, P. H., Taylor, M. M., & Forbes, S. M. (1968). Attention and multidimensional discrimination. *Perception & Psychophysics*, **4**, 113-117.

Lu, Z.-L., & Dosher, B. A. (1998). External noise distinguishes attention mechanisms. *Vision Research*, **38**, 1183-1198.

Lu, Z.-L., & Dosher, B. A. (1999). Characterizing human perceptual inefficiencies with equivalent internal noise. *Journal of the Optical Society of America A*, **16**, 764-778.

Luce, R. D., & Green, D. M. (1978). Two tests of a neural attention hypothesis for auditory psychophysics. *Perception & Psychophysics*, **23**, 363-371.

Lutfi, R. (1993). A model of auditory pattern analysis based on component-relative-entropy. *Journal of the Acoustical Society of America*, **94**, 748-758.

Massaro, D. W. (1975). Backward recognition masking. *Journal of the Acoustical Society of America*, **58**, 1059-1065.

Miller, J., & Bonnel, A.-M. (1994). Switching or sharing in dual-task line-length discrimination? *Perception & Psychophysics*, **56**, 431-446.

Milner, B., Taylor, L., & Sperry, R. W. (1968). Lateralized suppression of dichotically presented digits after commissural section in man. *Science*, **161**, 184-186.

Moore, J. J., & Massaro, D. W. (1973). Attention and processing capacity in auditory recognition. *Journal of Experimental Psychology*, **99**, 49-54.

Moray, N. (1970). Introductory experiments in auditory time sharing: Detection of intensity and frequency increments. *Journal of the Acoustical Society of America*, **47**, 1071-1073.

Narain, C., Scott, S. K., Wise, R. J., Rosen, S., Leff, A., Iversen, S. D., & Matthews, P. M. (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. *Cerebral Cortex*, **13**, 1362-1368.

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review*, **86**, 214-255.

Neff, D. L. (1995). Signal properties that reduces masking by simultaneous, random-frequency maskers. *Journal of the Acoustical Society of America*, **98**, 1909-1920.

Neff, D. L., & Green, D. M. (1987). Masking produced by spectral uncertainty with multicomponent maskers. *Perception & Psychophysics*, **41**, 409-415.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, **7**, 44-64.

Parkinson, S. R. (1974). Variability and control in dichotic memory. *Journal of Experimental Psychology*, **102**, 67-80.

Parkinson, S. R., Knight, M. V., DeMaio, J. C., & Connors, J. F. (1974). Channels and order of report in dichotic memory. *Perception & Psychophysics*, **16**, 235-241.

Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, **26**, 100-107.

Shiffrin, R. M., Pisoni, D. B., & Castaneda-Mendez, K. (1974). Is attention shared between the ears? *Cognitive Psychology*, **6**, 190-215.

Shinn-Cunningham, B., & Ihlefeld, A. (2004). Selective and divided attention: Extracting information from simultaneous sound sources. In S. Barrass & P. Vickers (Eds.), *Proceedings of ICAD '04: Tenth Meeting of the International Conference on Auditory Display*. Available online at www.icad.org/websiteV2.0/Conferences/ICAD2004/papers/shinn-cunningham_etal.pdf.

Sperling, G. (1960). The information available in brief presentations. *Psychological Monographs*, **74**, 1-29.

Sperling, G. (1984). A unified theory of attention and signal detection. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 103-181). Orlando: Academic Press.

STUDDERT-KENNEDY, M., & SHANKWEILER, D. (1970). Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America*, **48**, 579-594.

TERVANIEMI, M., & HUGDAHL, K. (2003). Lateralization of auditory-cortex functions. *Brain Research Review*, **43**, 231-246.

TREISMAN, A. [M.] (1964). Verbal cues, language and meaning in selective attention. *American Journal of Psychology*, **77**, 206-219.

TREISMAN, A. M. (1969). Strategies and models of selective attention. *Psychological Review*, **76**, 282-299.

TULVING, E., & LINDSAY, P. H. (1967). Identification of simultaneously presented simple visual and auditory stimuli. *Acta Psychologica*, **27**, 101-109.

VOGEL, E. J., & LUCK, S. J. (2002). Delayed working memory consolidation during the attentional blink. *Psychonomic Bulletin & Review*, **9**, 739-743.

WATSON, C. S., & KELLY, W. J. (1981). The role of stimulus uncertainty in discrimination of auditory patterns. In D. J. Getty & J. H. Howard, Jr. (Eds.), *Auditory and visual pattern recognition* (pp. 37-59). Hillsdale, NJ: Erlbaum.

WICHMANN, F. A., & HILL, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, **63**, 1293-1313.

WICKENS, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 63-102). Orlando: Academic Press.

WOOD, N., & COWAN, N. (1995). The cocktail party phenomenon revisited: Attention and memory in the classic selective listening procedure of Cherry (1953). *Journal of Experimental Psychology: General*, **124**, 243-262.

NOTE

1. Performance was so poor in Experiment 2 that identification was tested in a condition with no noise as well. This revealed that even with no noise present, identification performance was between 85% and 90% correct.

(Manuscript received November 11, 2005;
revision accepted for publication November 27, 2006.)