Attention, Perception, & Psychophysics
2010, 72 (2), 501-516
doi:10.3758/APP.72.2.501

Improved segregation of simultaneous talkers differentially affects perceptual and cognitive capacity demands for recognizing speech in competing speech

Alexander L. Francis
Purdue University, West Lafayette, Indiana

Perception of speech in competing speech is facilitated by spatial separation of the target and distracting speech, but this benefit may arise at either a perceptual or a cognitive level of processing. Load theory predicts different effects of perceptual and cognitive (working memory) load on selective attention in flanker task contexts, suggesting that this paradigm may be used to distinguish levels of interference. Two experiments examined interference from competing speech during a word recognition task under different perceptual and working memory loads in a dual-task paradigm. Listeners identified words produced by a talker of one gender while ignoring a talker of the other gender. Perceptual load was manipulated using a nonspeech response cue, with response conditional upon either one or two acoustic features (pitch and modulation). Memory load was manipulated with a secondary task consisting of one or six visually presented digits. In the first experiment, the target and distractor were presented at different virtual locations (0º and 90º, respectively), whereas in the second, all the stimuli were presented from the same apparent location. Results suggest that spatial cues improve resistance to distraction in part by reducing working memory demand.

Early studies of selective attention were conducted almost exclusively in an auditory domain, focusing largely on the problem of how listeners were able to direct attention to the speech of one talker when presented simultaneously with that of one or more additional talkers, a problem dubbed the *cocktail party problem* by Cherry (1953). The results of these studies suggested that perception was best understood as a mechanism with limited capacity for information processing, although whether the limitation occurred early in processing (e.g., Treisman, 1964, 1969) or later (e.g., Deutsch & Deutsch, 1963) was a matter of some debate (see Driver, 2001). Although more recent research on the perception of speech in competing speech has not generally referenced current theories of selective attention, the assumption that speech perception depends on the operation of limited capacity mechanisms is widespread and is entirely compatible with current cognitive neuroscientific perspectives on the role of resource limitations in governing information processing.

Despite the early correspondence between attentional theory and research on perception of speech in competing speech, more recent research on the role of capacity limitations in perception has progressed mainly in the visual domain (e.g., Lavie, 2005). In contrast, investigations of the mechanisms that solve the cocktail party problem in human listeners have been pursued primarily in the context of audiological research (e.g., Humes, 2002), often under the rubric of *informational masking* with little reference to mechanisms of attention (see the discussion by Shinn-Cunningham, 2008), and/or in studies of cognitive and perceptual aging (Pichora-Fuller, Schneider, & Daneman, 1995; Tun, O'Kane, & Wingfield, 2002). The majority of these speech-oriented studies have focused strongly on the role of sensory impairment (e.g., Humes, 2002) and/or acoustic properties of the signal (e.g., Vongpaisal & Pichora-Fuller, 2007), and the emphasis has often been on finding a peripheral auditory explanation for older listeners' disproportionate difficulty in understanding speech in competing speech (see the discussion by Humes, Lee, & Coughlin, 2006). Even when the theoretical focus has been on more central cognitive factors (e.g., Wingfield & Grossman, 2006), until recently there has typically been little emphasis on current theories of selective attention per se (although, for representative articles making explicit reference to demands on limited capacity within the framework of informational-masking research, see Gallun, Mason, & Kidd, 2007; Kidd, Mason, Richards, Gallun, & Durlach, 2007; Yost, 2006).

Within the field of informational masking, one consistent finding is that spatial separation of the target and distracting speech facilitates perception, in a phenomenon often referred to as *spatial release from masking* (SRM). Some, but not all, of this release is due to a release from energetic masking (the physical interfer-

A. L. Francis, francisa@purdue.edu

ence of one signal with another). For example, Freyman, Helfer, McCall, and Clifton (1999) found that a spatial separation between target speech and masking speech of 60º reduced masking thresholds by about 13.7 dB, as compared with a reduction of only 8.2 dB when the masker was speech-shaped noise. This suggests that spatial separation provides an additional benefit (of about 5.5 dB in this case) over that of spatial release from energetic masking when conditions that allow for informational masking are dealt with (see also qualitatively similar findings, using a different method, in the work of Chan, Merrifield, & Spence, 2005). Further evidence for this hypothesis was shown by Freyman, Balakrishnan, and Helfer (2001), who found a decrease in masking threshold of about 4–6 dB when an English-speaking masker was perceived to be separated from an English target by 60º but of only 3–4 dB when the same masking talker was speaking Dutch (and was, therefore, unintelligible to the English-speaking listeners, reducing overall informational-masking levels). Freyman et al. (1999) proposed that perceived spatial separation provides additional cues that facilitate auditory scene analysis (formation of distinct auditory objects), even when the separation is completely illusory (cf. Driver, 1996).

The present study is concerned with the consequences of such facilitation for speech perception within the framework provided by load theory (Lavie, 2000, 2005), a recent theory of selective attention based mainly on evidence from visual studies. Applying a more general theory of selective attention to the study of speech perception accords well with the suggestions of Shinn-Cunningham (2008), and the emphasis of this specific theory on distinguishing between levels of attentional processing recommends it for investigating informational masking, a phenomenon that clearly involves multiple levels at which distractors may interfere with target processing.

According to load theory, selective attention is governed by two kinds of capacity-limited processes. At the perceptual level, limited capacity serves to reduce the intrusion of irrelevant distractors in a passive manner, in the sense that stimuli that do not receive sufficient attention are not processed, resulting in early selection (early exclusion of irrelevant objects before they are fully processed perceptually). When there is surplus attentional capacity available beyond what is necessary for performing required tasks (processing intended targets and doing whatever else might be demanded, e.g., in a dual-task situation), this remaining capacity is obligatorily devoted to processing whatever else is in the scene, including irrelevant distractors, which may therefore interfere with subsequent processing of target stimuli. When the primary task(s) demand more perceptual attention, the surplus capacity available for processing irrelevant distractors decreases, and distractors interfere less.

In contrast, at the cognitive level, load theory predicts that decreasing the availability of surplus capacity will lead to increased interference from irrelevant distractors, because, here, capacity-limited cognitive mechanisms filter out irrelevant distractors that have intruded at the perceptual level, resulting in late selection (the exclusion of objects from further processing only after significant analysis has already been performed upon them). At this level, capacity is considered in terms of working memory (WM), in the sense that WM is necessary to maintain task goals (de Fockert, Rees, Frith, & Lavie, 2001), or, more recently, in terms of executive control processes, in the sense that the load cost is associated with the process of switching between tasks—for example, a secondary WM task and the primary perceptual task (Brand-D'Abrescia & Lavie, 2008)—or both (Lavie, Hirst, de Fockert, & Viding, 2004). In either case, imposing additional cognitive load on the capacity-limited system leads to a decrease in the ability to filter out irrelevant stimuli.

Because of its emphasis on the role of capacity limitations at different levels of processing, load theory provides an excellent framework for exploring the consequences of spatial separation on speech perception in competing speech. Recent research has shown that some aspects of informational masking are capacity limited, whereas others are not (Alain, Reinke, He, Wang, & Lobaugh, 2005), suggesting that it may be possible to identify the level(s) at which a particular manipulation (i.e., spatial separation) influences informational masking by identifying the interaction of that manipulation with capacity demands at particular levels of processing.

One of the major criticisms of resource theories, including load theory, is that it is difficult to quantify the *load* demand of a particular task in a purely behavioral fashion (e.g., Allport, 1993). However, functional imaging studies of brain activity during the performance of tasks varying in difficulty suggest a correspondence between task difficulty (i.e., effort) and cortical activity in visual (Carpenter, Just, Keller, Eddy, & Thulborn, 1999), language (Keller, Carpenter, & Just, 2001), and speech (Wong, Nusbaum, & Small, 2004) tasks. Furthermore, studies by Lavie and colleagues have shown a modulation of distractor-related cortical activity as a function of cognitive (de Fockert et al., 2001) and perceptual (Rees, Frith, & Lavie, 1997) load imposed by a secondary task in dual-task paradigms. Thus, although the operational definition of *load* remains vague, it seems clear that, at a neurological level, increased processing demand is related to increased cellular activity in specific brain regions and that increasing the processing demand incurred by one task can significantly reduce cortical activity related to performance of another task in a manner compatible with the basic principles of resource theories.

Although the potential benefits of applying load theory to speech perception are clear, the results of a recent study by Gomes, Barrett, Duff, Barnhardt, and Ritter (2008) suggest that auditory processing may not be subject to the same kinds of capacity limitations as visual attention, especially with respect to the effects of perceptual load. In their study, listeners were asked to monitor one of two pitch-defined channels (e.g., low pitch = 1000 Hz or high = 2000 Hz) for the appearance of lower intensity tokens (76-dB SPL targets among 82-dB SPL standards) while ignoring the other channel, which also contained both low- and high-intensity tokens. Gomes et al. manipulated the perceptual load of this task by changing interstimulus intervals (ISIs; 300, 600, or 900 msec for high, medium, and low load) and measured the onset latency and amplitude of a variety of evoked response potentials, including the Nd wave, which they argued is indicative of a distinction between processing of attended and unattended stimuli. Their results showed no evidence, either behaviorally or electrophysiologically, of increased processing of distractor tokens under low perceptual load (longer ISI), suggesting that the assumption of load theory (specifically, that excess perceptual capacity must be obligatorily allocated) may not be applicable in an auditory domain, although their results still support the general hypothesis that auditory attention involves the operation of at least one limited capacity system. This suggests that the predictions of load theory, at least with respect to the effects of perceptual load, may not be supportable in an auditory modality.

With respect to cognitive (including WM) load, there is already considerable evidence to support the hypothesis that cognitive capacity limitations play a significant role in speech perception. For example, Nusbaum and colleagues (Nusbaum & Magnuson, 1997; Nusbaum & Morin, 1992; Wong et al., 2004) have shown that listeners are significantly slower to recognize speech when it is presented in a context in which the talker is variable than when all the speech they hear is produced by a single talker. This increased response time (RT) interacts with WM load (Nusbaum & Morin, 1992), suggesting that listening to speech from multiple talkers, even when they are presented sequentially, requires the commitment of WM resources. Nusbaum and colleagues suggested that this increased memory load is related to the need to maintain a greater number and/or variety of acoustic signal properties (spectral/temporal and spatial) in memory when talkers are varying, both because linguistically relevant cues vary in a talker-dependent manner (cf. Nusbaum & Magnuson, 1997) and because talker differences are typically cued by acoustic and spatial differences (see Wong et al., 2004, for a discussion and neural evidence supporting this hypothesis).

The interaction of cognitive and perceptual resources is of particular interest when the specific problem of listening to a single talker in the presence of irrelevant speech or other noise is considered. In this case, a number of researchers (e.g., Pichora-Fuller et al., 1995; Rabbitt, 1991; Tun et al., 2002) have proposed that the age-related increase in word recognition difficulty derives from the need to reallocate limited cognitive resources, diverting capacity from higher level cognitive systems, such as memory encoding or message comprehension, to lower level systems involved in perception. For example, Pichora-Fuller et al. showed that older listeners were worse than younger listeners at recalling words that they had already identified after hearing them in sentences presented at a variety of signal-to-noise ratios (S/N). Moreover, even for younger listeners, recall accuracy for previously identified words decreased as S/N decreased. Pichora-Fuller and colleagues argued that the poorer hearing of the older listeners required them to devote a larger proportion of a limited pool of resources to auditory processing, as did the decreasing S/N for both older and younger listeners, leaving fewer resources available for memory encoding (cf. Rabbitt, 1968, 1991, for similar findings and arguments).

Thus, a variety of studies support the hypothesis that speech perception depends on the operation of at least one capacity-limited system, probably involving WM and/or executive control, but possibly also another earlier, more sensory-specific system for which the commitment of higher level cognitive resources may be able to compensate to some degree—for example, in cases of lower S/N or hearing impairment. On the other hand, although most theories of spatial release from masking currently favor explanations in terms of changes in perceptual processing (i.e., auditory scene analysis), it is not clear whether auditory object processing is subject to the same kind of capacity limitations as visual object processing, and, in addition, much of the benefit of spatial release from informational masking seems to accrue at a more cognitive level, as is suggested by the interaction of SRM and the informational content of the masking stimulus. Therefore, on the basis of existing evidence, it is not clear whether the effects of spatial separation of target and masker interact with capacity demands at either a perceptual or a cognitive level. In keeping with the suggestion of Shinn-Cunningham (2008), in the present study, this question was investigated by applying the predictions of load theory, a general theory of attention, within an auditory (speech) modality, with the aim of better understanding the consequences of spatial separation on demands for WM and perceptual capacity when speech is listened to in the presence of competing speech.

In two experiments, listeners performed a primary auditory flanker task based on that used by Chan et al. (2005) under different levels of both perceptual and cognitive load. In this task, listeners were asked to attend to a single voice with a distractor voice (flanker) of the opposite gender. The two voices could produce either the same word (congruent condition) or different words (incongruent condition). In the first experiment, the target voice was presented diotically (thus, appearing to be located at the midline or in the center of the listener's head), the distractor was presented monaurally (thus, appearing to be located directly to one side of the listener), and the target was identified spatially (i.e., "listen to the center voice"). In the second experiment, the target and distractor were both presented diotically, and the target was identified by gender (i.e., "listen to the male voice"). In both experiments, the primary task was always to identify the word produced by the target talker. Interference was calculated as the difference between the RT on incongruent trials and that on congruent trials. Load theory predicts that listeners should show decreasing interference as perceptual load increases and increased interference as cognitive (WM) load increases (Lavie, 2000, 2005).

EXPERIMENT 1

The goal of the first experiment was primarily to determine the applicability of load theory to perception of speech in competing speech in an auditory paradigm maximally comparable to the typical visual flanker paradigm used in previous studies of load theory. In this experiment, the effects of perceptual load were compared with those of WM load in an auditory flanker task in which the target and distractor were clearly spatially distinct. Perceptual load was manipulated by increasing the processing demands of a secondary response cue in a manner comparable to that in the work of Lavie and colleagues (Lavie, 1995; see also Alain & Izenberg, 2003) and consistent with feature integration theory (Treisman & Gelade, 1980; Treisman & Sato, 1990; see Lavie, 1995, for a discussion): To induce high load, participants were asked to attend to two features of the response cue (pitch and modulation), making a response only on trials in which the cue had a specific combination of properties of both features—for example, *low pitch and modulation* or *high pitch and no modulation*. Low load was imposed by asking the listeners to respond only when the cue exhibited a particular value of a single feature (e.g., *modulated*). Perceptual load was manipulated by means of an auditory cue because previous research (e.g., Rees, Frith, & Lavie, 2001) had shown that perceptual load effects are strongest (or perhaps, appear only) when the secondary task is presented in the same modality as the primary one.

Cognitive load was manipulated using a secondary digit memory load identical to that used by Lavie et al. (2004; see also Francis & Nusbaum, 2009; Nusbaum & Morin, 1992; Wong et al., 2004). Memory load was imposed using visually presented digits, rather than auditory ones, because this is the most common method for introducing working memory load and using this method maximized continuity with previous studies related to load theory. Moreover, a fundamental assumption of load theory is that interference arising at the cognitive level does so in a modality-independent manner because it results from capacity limitations on executive control, rather than on sensory-specific processes (Brand-D'Abrescia & Lavie, 2008). Therefore, the modality of the memory load stimuli was expected to be irrelevant. Following the discussion in Lavie (2000, 2005), interference was expected to increase as WM load increased but to decrease as perceptual load increased.

Method

Participants

Forty-three participants (15 men, 28 women) between the ages of 18 and 28 years (mean, 21 years, 2 months) were recruited from the campus of Purdue University for this study. Of these, 4 served as pilot participants, and so their data were not included in the final total; 7 were excluded from further analysis because they failed to meet threshold criteria for normal hearing (1) or right-handedness (4) or were unable to complete the task (2). All the remaining participants ($N = 32$) were right-handed native speakers of a North American dialect of English with no significant foreign language experience and no history of attention deficit, all by self-report. All demonstrated hearing within normal limits bilaterally, as determined by pure tones presented at 25 dB HL at 500, 1000, 2000, and 4000 Hz through a GSI-61 portable audiometer. All testing was conducted under a protocol approved by the Committee for the Protection of Human Research Subjects at Purdue University, and all the participants were paid for their time.

Stimuli

Auditory stimuli consisted of stereo .wav files presenting a combination of two talkers and (in the perceptual load conditions) a nonspeech amplitude-modulated tone. The stimuli were constructed according to the following procedures. Multiple productions of the words *bead* and *bad* were recorded by one male and one female native speaker of English. Recordings were conducted in a double-walled sound booth (IAC Inc.) in the audiology clinic at Purdue University, using a Marantz PMD 660 digital audio recorder and a hypercardioid microphone (Audio-Technica D1000HE) positioned on a boom approximately 20 cm in front and 45º to the right of the talker's lips. Recordings were made at a sampling rate of 44.1 kHz with a 16-bit quantization rate (QR) and were subsequently peak-amplitude normalized to the maximum QR. Durations were measured, and four tokens were selected, one for each word produced by each talker. Tokens were selected so as to match closely in duration, subject to the requirement that they also be free of acoustic artifacts, both vocal (i.e., breathy noise and glottalization) and extrinsic to speech (e.g., electrical clicks, paper rustling, etc.). All the tokens were produced without a final burst release, and durations ranged from 466 msec (male *bead*) to 522 msec (female *bad*). All the stimuli were subsequently normalized in duration (520 msec) using PSOLA resynthesis implemented in Praat 4.6 (Boersma & Weenink, 2007). All the stimuli retained their individual, naturally falling fundamental frequency ($f0$) contour. Exact $f0$ properties are shown in Table 1. These speech tokens were then combined into stereo sound files by writing an identical version of the target sound file to each channel at half the original amplitude and subsequently adding one full-amplitude version of the distractor sound file to either the left or the right channel. This means that the target-to-distractor intensity ratio was approximately equal, and, although presenting the target signal to both ears while the distractor was presented only to one may have reduced masking by the distractor to some degree, due to binaural masking level difference (MLD) effects (Moore, 2003), all the stimuli were presented well above threshold, making energetic masking less of a concern in this case. The gender of the talker producing the distractor was always the opposite of that of the talker producing the target, but the target and distractor could be either the same word or different ones. In this way, a total of 16 stimuli were generated (2 target talkers × 2 target words × 2 distractor words × 2 flanker ears).

For the stimuli used in the perceptual load task, an additional signal was added to each stereo file to serve as a response cue. This signal was 520 msec long and consisted of either a low (1300 Hz) or a high (3900 Hz) pure tone that was either sinusoidally amplitude modulated at approximately 18 Hz (exact modulation frequency was chosen so that the onset and offset of the signal coincided with zero-crossings of the modulation sinusoid), with a modulation depth of 50%, or not modulated at all. Frequency and modulation values were selected to maximize the perceptual distinctiveness of each signal and, in the case of frequency, to minimize energetic masking between the response cue and the first three formants of all four speech sounds (see Figure 1). Amplitude ramped up linearly from 0 to maximum over the first 50 msec and went back down again over the final 50 msec.[1] After the four signals (low and high pitches with and without modulation) were generated, they were normalized to be equivalent in RMS amplitude (with a peak amplitude of approximately 48% of the total QR for the unmodulated signals and 67%

Table 1
Speech Stimulus Parameters

Word	Talker	Onset $f0$ (Hz)	Offset $f0$ (Hz)	Average $f0$ (Hz)
Bead	Female	257	181	217
	Male	119	88	121
Bad	Female	257	181	212
	Male	135	103	107

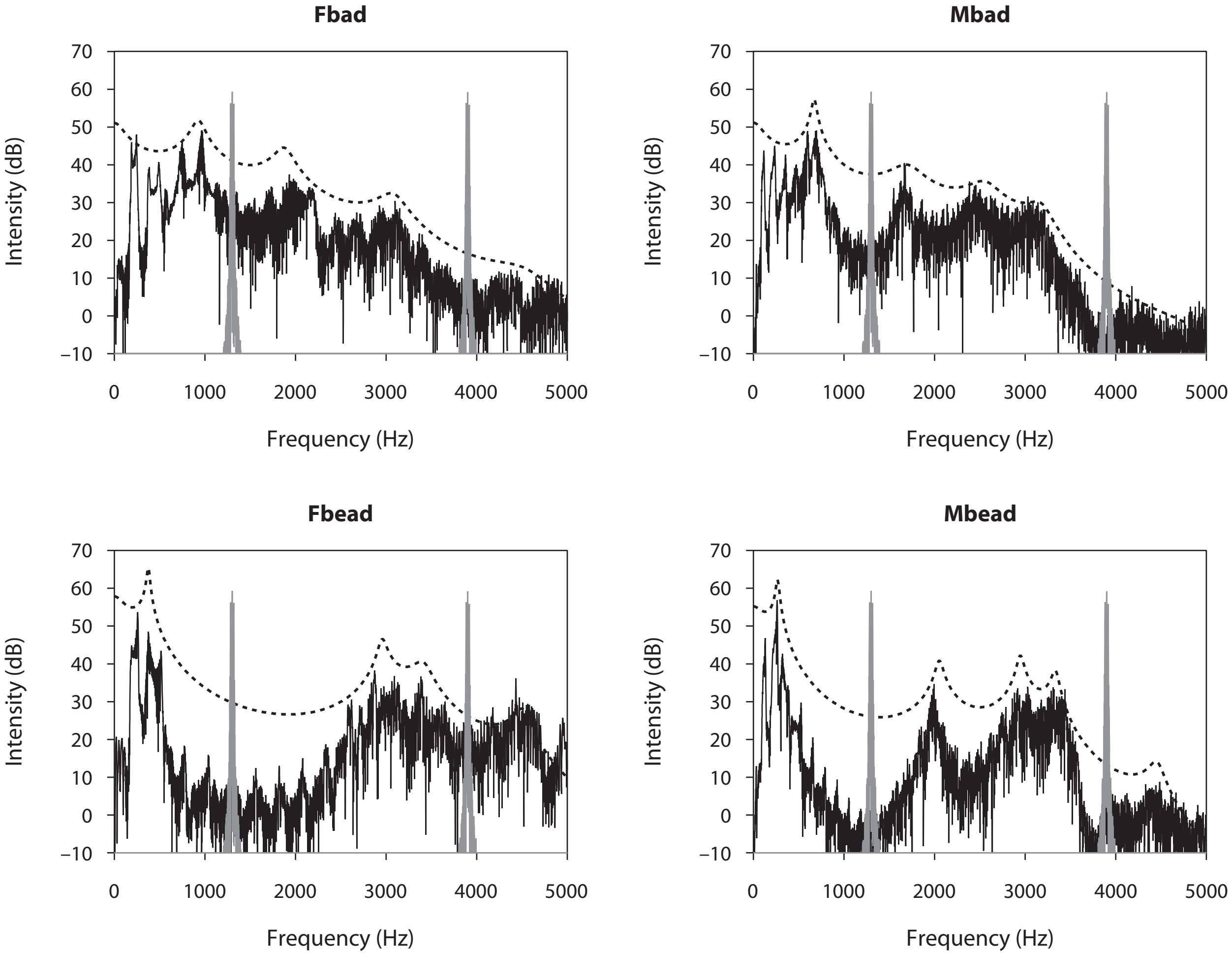

Figure 1A. Stimuli for both experiments: Spectral distribution of energy for each of four speech stimuli with each of two different response cues (modulated tones with 1200- and 3900-Hz center frequency, respectively). Response cue energy is distributed symmetrically around the center frequency and includes sidebands resulting from modulation. Unmodulated response cues would appear as a single line centered within the representation of the modulated tone spectrum shown here. Note that the 1200-Hz response cue lies between the first formant (*F*1) and *F*2 of all four spectra, whereas the 3900-Hz response cue lies above *F*4 for the male talker (right column) and between *F*3 and *F*4 for the female talker (left column). Mbad, male talker, saying *bad*; Mbead, male talker, saying *bead*; Fbad, female talker, saying *bad*; Fbead, female talker, saying *bead*. Intensity values (listed in decibels) are defined with respect to an arbitrary reference but are consistent across all stimuli.

for the modulated ones). Each signal was added to both channels of each of the 16 speech stimuli, resulting in a total of 64 stimuli in which the response cue was heard as coming from the same midline location as the target voice stimulus at approximately half the peak intensity of the target (although both the cue and target were significantly above threshold and clearly audible). Sixteen vocal stimuli without the response cue were also retained for familiarization and memory load testing.

Procedure

The experiment was run on a Dell Optiplex running Windows XP using E-Prime version 1.2.1.841 (E-Run 1.2.1.61; Schneider, Eschman, & Zuccolotto, 2002). The stimuli were presented via Sennheiser HD-25 headphones driven by a Soundblaster Live! sound card. All the visual stimuli were presented in black type against a light gray background on a 15-in. LCD screen. Responses were collected using a Cedrus RB620 six-button response pad that the participants were instructed to hold in a comfortable manner, with one finger or thumb on each of the leftmost and rightmost buttons. Response accuracy and RTs were recorded.

The participants completed four blocks of trials, two with perceptual load (one low, one high) and two with memory load (one low, one high), in about 60 min. On all the trials, the task was the same: Identify the word (either *bead* or *bad*) spoken by the central talker and ignore the peripheral one. Task order (memory load vs. perceptual load) and load order (high–low or low–high) were counterbalanced across listeners, except that load order for a given participant was maintained across tasks, so that, for example, a participant who completed the low-memory-load block before the high-memory-load block also completed the low-perceptual-load block before the high one. The ear to which the flanker stimulus was presented and the left–right order of response choices displayed on the screen (*bead* vs. *bad*) and on the corresponding response pad buttons were independently counterbalanced across listeners in such a way that a participant heard the flanker in the same ear on all the trials and always saw either *bead* on the left and *bad* on the right (corresponding to a left buttonpress to indicate having heard *bead* and a right one for *bad*) or vice versa.

Before beginning the test blocks, the participants were presented with text describing an overview of the entire experiment. Then they

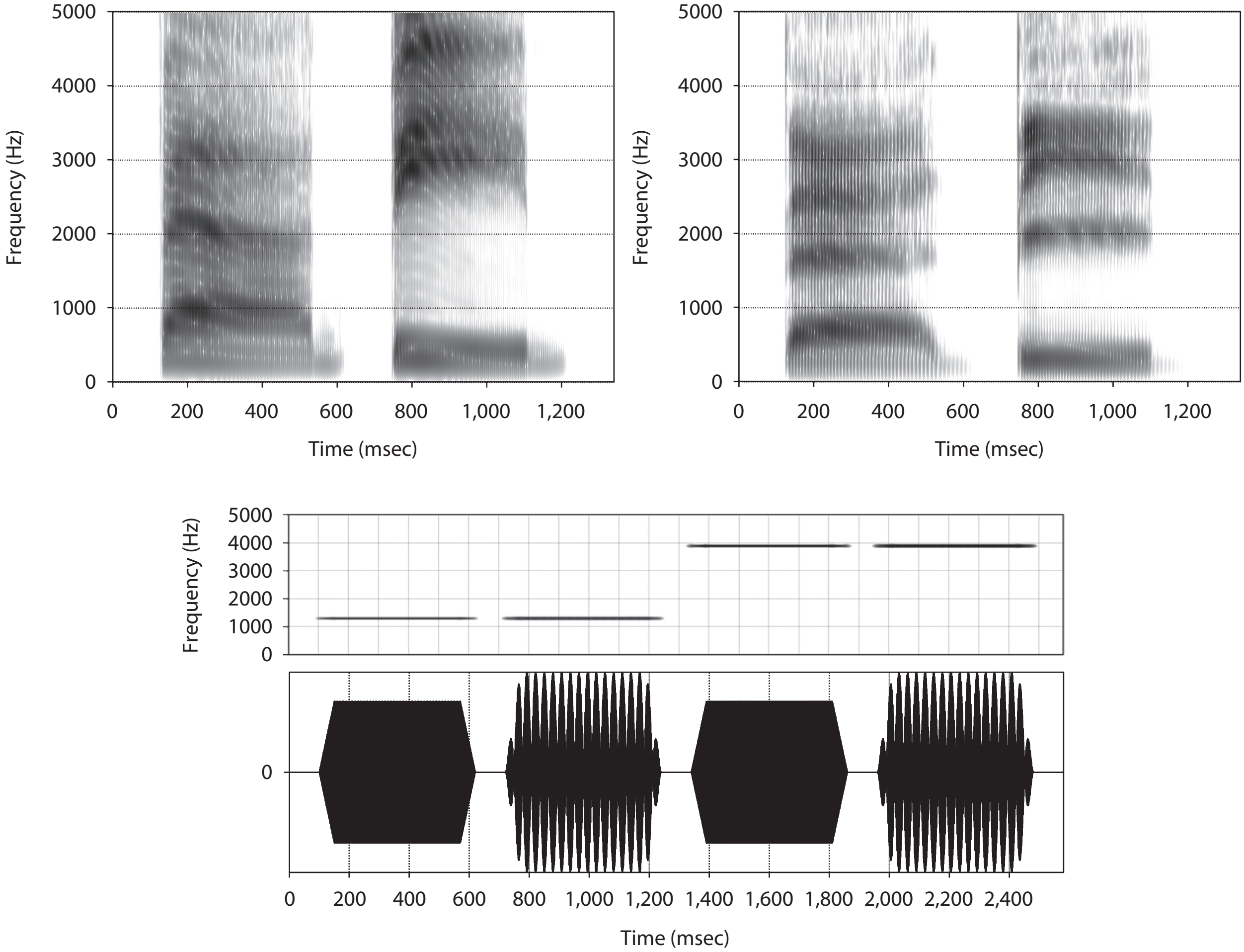

Figure 1B. Stimuli for both experiments. Bottom panel: Acoustic waveforms and spectrographic displays for the four response cues: low and unmodulated, low and modulated, high and unmodulated, and high and modulated. Top left: Spectrographic displays of female talker's productions of *bad* and *bead*. Top right: Spectrographic displays of male talker's productions of *bad* and *bead*.

were given a set of 24 flanker task trials consisting of three repetitions of each of eight tokens (2 target talkers × 2 target words × 2 distractor words). Feedback was provided following each trial. The participants who reached a criterion of 80% correct responses at any point after the first 8 trials immediately continued on to the main study. Those who did not reach criterion in the first 24 trials stopped the experiment and had the experimental task explained to them again, after which they were allowed to start the experiment once more from the beginning. If they were still unable to reach criterion on the second attempt, they were dropped from the experiment. After training was complete, the participants began either the memory or the perceptual load task.

Perceptual load. The participants completed two perceptual load blocks, one under low load and one under high. Imposition of low and high perceptual load was accomplished by requiring the listeners to attend to either a single feature (modulation) or two features (pitch and modulation), respectively (see Lavie, 1995, and Treisman, 1964, for arguments that attending to more than one feature incurs greater capacity demands). This was implemented by asking the listeners to respond to the primary task only if the response cue exhibited requisite pitch and modulation properties. For low-load trials, the listeners had to attend to only one feature (modulation)—for example, "Respond only if the tone is *trilled*." On high-load trials, the listeners were asked to respond on the basis of two features of the cue—that is, "Only respond if the tone is either *low and trilled* or *high and smooth*).

Before starting the first perceptual load block, the participants saw a brief description of the task and went through a brief, self-paced auditory presentation and written description of the four possible response cues. Each load block was also preceded by a set of 32 training trials (24 *go* trials and 8 *no-go*) at the appropriate level of load, with trial-level feedback on both the flanker task performance (whether the participant reported hearing the correct word) and the response cue (the modulation and pitch properties of the presented cue and whether those meant that the trial was a go or a no-go trial). The participants were given 2,520 msec to respond from the start of the stimulus. Performance on this training set was not analyzed, and all the participants went on to the test blocks regardless of score.

Each perceptual load block consisted of 128 experiment trials (96 go and 32 no-go). Perceptual load trials had the same structure in both the low- and high-load blocks. The trial began with presentation of the response choices, *bead* and *bad*, in 18-point serif font (Courier), vertically centered and spaced at approximately 25% of the screen from the left and right sides. A reminder of the response cue (e.g., for low load, "Respond only if the tone is *trilled*," and for high load, "Respond only if the tone is *low and trilled* or *high and smooth*") was shown at the bottom of the screen for every trial.[2] All the text appeared on the screen 1 sec before the auditory stimulus (520 msec) began and remained until a response was made or the trial ended 2,520 msec after the start of the auditory stimulus. Once a response was made, the next trial began immediately with a 1-sec presentation of the response choices and response cue reminder text.

Memory load. The participants also completed two memory load blocks, one with low and one with high memory load. A low memory load was imposed by asking the listeners to remember a single digit before beginning the flanker task, whereas high-memory-load trials required remembering a list of six digits.

Before starting the memory load blocks, the participants first saw a general description of the task. The low- and high-memory blocks then each started with a more specific explanation of the task—in particular, identifying the number of digits to be remembered on each trial—followed by 16 practice trials, each having exactly the same structure as the respective experiment trials, but with trial-level feedback on both flanker and memory task performance.

The structure of the memory load trials was identical across blocks, except that more time was given to encode the stimuli in the high-load trial (see Figure 2). On each memory load trial, the participants first saw a fixation cross in the center of the screen (500 msec), followed by either one (low load) or six (high load) digits centered on the screen. The digits were presented in 24-point, serif (Courier) font, with two spaces between each pair of digits on the high-load trials. Digits were randomly selected, with the constraint that no more than two of the same digit could appear in a given list. In the low-load condition, the digits were presented for 500 msec; in the high-load condition, they remained on the screen for 2 sec. Following digit presentation, a visual mask consisting of a # symbol in the same font and at the same location as each digit in the list was presented for 750 msec in the low-load condition and 2,500 msec in the high. The flanker task proceeded in the same manner as in the perceptual load blocks, except that there was no text referring to the response cue shown on the screen, since there was no response cue present in the auditory stimulus. Following a response to the flanker task or 2,520 msec after the onset of the auditory stimulus, the participants were shown a single number and were asked to press the left button if that number was in the list they were remembering and the right button if it was not. Once a response was made or after 5 sec, the trial ended and the next trial began with a fixation cross.

In both memory and perceptual load conditions, under both high and low load, accuracy and RTs on correct responses were recorded. RTs greater than three standard deviations outside the mean were excluded from further analyses on a by participant, by condition basis (approximately 1% of RTs were thus excluded). RT analyses were conducted only on trials in which the participants were correct on both the primary and secondary tasks, in order to ensure that these results truly reflected accurate performance under the intended load level.

Results

RTs and interference scores for both the perceptual and memory load tasks are shown in Figure 3. Although the experimental design was fully crossed and within subjects, there were no specific predictions regarding a quantitative relationship between performance in the memory and perceptual load conditions, so the results from the two tasks were analyzed separately.

Perceptual Load

In the perceptual load task, performance on the secondary task was scored in terms of the signal detection statistic d', calculated in terms of z(hits) $-$ z(false alarms), where a hit consisted of a response in a go condition, whether the response was correct with respect to the primary task or not, and a false alarm consisted of a response on a no-go trial, whether the response was correct or not. A two-way repeated measures ANOVA showed significant effects of load level [$F(1,31) = 70.68, p < .001$], but not of trial type [$F(1,31) = 1.05, p = .31$], and no interaction [$F(1,31) = 0.04, p = .84$], with listeners averaging a d' of

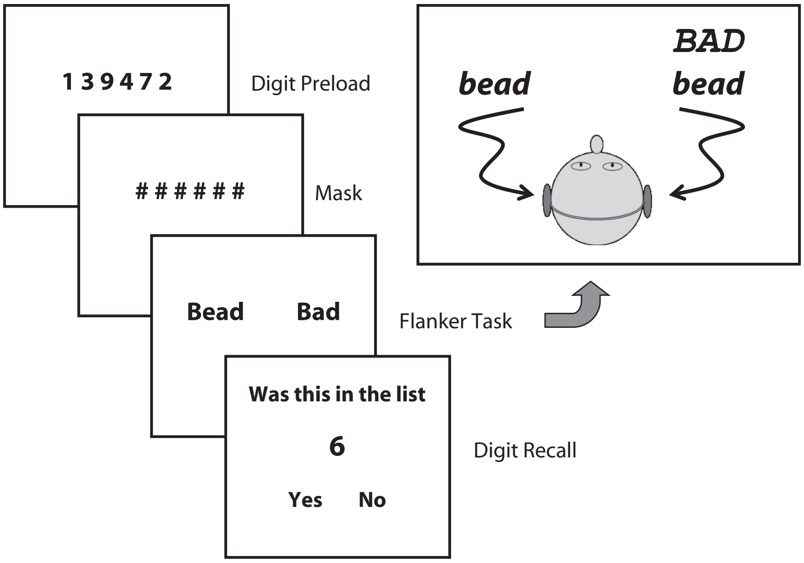

Figure 2. Schematic diagram of a single trial in the high-memory-load condition in Experiment 1, illustrating digit preload, mask, flanker task, and recall task. Inset shows schematic diagram of stimulus presentations during an incongruent trial of the flanker task, showing talker locations at 0º (target, sans serif font) and 90º (distractor, serif font) simulated by presenting the target diotically and the distractor monaurally, as in Experiment 1. Note that perceptual load trials also presented a nonspeech response cue at 0º (see the text).

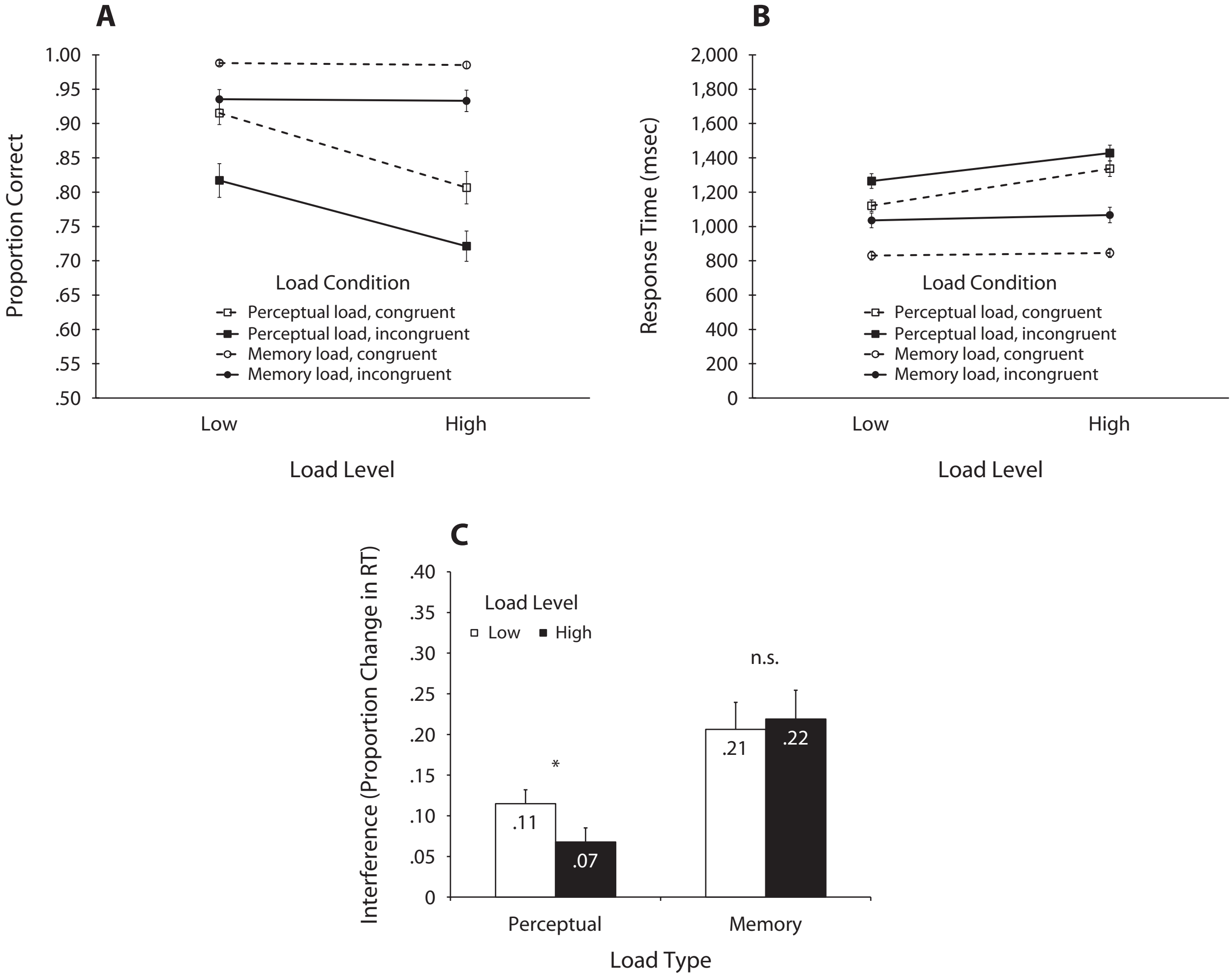

Figure 3. Results for Experiment 1, showing proportions of correct responses (A), response times (RTs) on correct responses for both congruent and incongruent trials of the auditory flanker task under low and high perceptual and memory load (B), and interference (proportion of increase in RT from congruent to incongruent conditions) under low and high perceptual and memory load (C). Error bars indicate 1 *SE*.

2.83 under low load and 1.43 under high load. Although the participants did have more difficulty with the high-load task, this did not interact with the congruency of the primary task trial.

On the flanker task, the participants were relatively accurate at recognizing the target word, averaging 86.1% correct in the congruent condition (91.5% in the low-load blocks, 80.7% in the high) and 76.9% correct in the incongruent condition (81.7% in the low-load blocks, 72.1% in the high). Note that these scores are only for trials in which a response was expected (i.e., go trials). A two-way repeated measures ANOVA showed a significant effect of trial type [$F(1,31) = 17.48$, $p < .001$] and of load level [$F(1,31) = 21.67$, $p < .001$] but no interaction between the two [$F(1,31) = 0.09$, $p = .77$]. Although the participants became slightly less accurate at recognizing the target word under high load and on incongruent trials, the effect of load did not differ between congruent and incongruent trials at either high or low load levels.

With respect to RTs, listeners were faster on congruent trials ($M = 1{,}229$ msec; 1,121 msec in low-load trials, 1,337 in high) than on incongruent ones ($M = 1{,}347$ msec; 1,265 msec in low-load blocks, 1,428 msec in high), and a two-way repeated measures ANOVA showed significant main effects of trial type [$F(1,31) = 7.66$, $p = .01$] and of load level [$F(1,31) = 19.99$, $p < .001$] but no interaction between the two [$F(1,31) = 0.39$, $p = .54$]. However, because load theory makes specific predictions about the interaction between trial type (congruent vs. incongruent) and load level (low vs. high) in terms of *interference*, quantified as the difference between incongruent and congruent RTs on correct responses, a Student's t test of dependent samples was conducted to compare interference in the low-load condition (144 msec) with that in the high (91 msec). The results were significant [$t(31) = 2.40$, $p = .023$], demonstrating that, as predicted by load theory, increased perceptual load results in a decrease in interference in a speech perception task. This finding is preserved even when secondary task performance, in terms of d' under high load, is included as a covariate in an ANCOVA [$F(1,31) = 6.35$, $p = .017$; covariate, $F(1,30) = 1.46$, $p = .24$], suggesting that individual variation in the ability to perform the high-perceptual-load secondary task did not affect ability to perform the primary flanker task. It should be noted, however, that RTs tend to show considerable variability across individuals, which may adversely affect within-group variance when differences between RT values are compared. For example, a listener who is generally slow to respond may exhibit an RT in a congruent condition of 800 msec, increasing to 1,600 msec on incongruent trials (a difference of 800 msec). In contrast, a listener who is relatively quick to respond might produce an average congruent RT of 400 msec and an incongruent RT of 800 msec (a difference of 400 msec). Although both listeners take twice as long to respond in the incongruent condition, as compared with the congruent one, and therefore may be judged to be equally affected by that manipulation, their *difference* scores are very different. Because of the inherent individuality of RTs, some researchers recommend comparing ratios rather than raw differences (e.g., Madden, 2001) or otherwise normalizing RT values to individual performance. Similarly, Lavie et al. (2004) compared raw RT differences for each subject (i.e., incongruent − congruent), as well as differences normalized against the average for the load condition [i.e., (incongruent − congruent)/mean(congruent, incongruent)], expressed as percentages of mean RT, in order to control for the fact that RTs in the high-load condition were uniformly higher than those in the low. In the present case, when interference was calculated as proportions of mean RT for the condition, the observed effect of load (low, 11.5%; high, 6.7%) remained significant, both when analyzed by a t test [$t(31) = 2.56$, $p = .016$] and when analyzed by an ANCOVA including d' as a covariate [$F(1,31) = 6.54$, $p = .016$; covariate, $F(1,30) = 1.42$, $p = .24$]. Thus, increasing perceptual load clearly resulted in a significant *decrease* in interference from an irrelevant talker, precisely as predicted by load theory.

Working Memory Load

In the memory load blocks, the listeners were quite accurate on secondary task performance (number recall). The results of a two-way repeated measures ANOVA showed a significant effect of load level [$F(1,31) = 15.27$, $p < .001$] but no effect of trial type [$F(1,31) = 3.28$, $p = .08$] and no interaction between the two [$F(1,31) = 0.05$, $p = .83$], with a mean score of 88.5% correct under high load (six digits) and 93.0% correct under low load (one digit).

On the primary task, the listeners were also very successful at recognizing target words. Response accuracy was 98.7% correct on congruent trials (98.8% correct in low-load blocks, 98.5% in high) and 93.4% correct on incongruent ones (93.6% in low-load blocks, 93.3% in high). A two-way repeated measures ANOVA of these scores showed a significant effect of trial type [$F(1,31) = 22.94$, $p < .001$], but not of load [$F(1,31) = 0.06$, $p = .81$], and no interaction [$F(1,31) < 0.01$, $p = .98$]. Although the participants were less accurate at recognizing the target word on incongruent trials, this difference was not affected by changes in load.

The effect of memory load on RT on correct trials in the flanker task was analyzed with a two-way repeated measures ANOVA with two levels of each factor: trial type (congruent or incongruent) and load level (low or high). The results showed a significant main effect of trial type [$F(1,31) = 34.53$, $p < .001$], with congruent trials ($M =$ 838 msec; low = 830 msec, high = 846 msec) being significantly faster than incongruent ones ($M = 1{,}051$ msec; low = 1,036 msec, high = 1,067 msec). The effect of load was not significant [$F(1,31) = 0.39$, $p = .54$], nor was the interaction [$F(1,31) = 0.09$, $p = .77$]. Again, because load theory makes specific predictions regarding the effect of changing working memory load on the difference between congruent and incongruent RTs (*interference*), a paired-comparison Student's t test was conducted on the interference (difference) scores. The results showed no significant difference between interference under low (205 msec) and high (221 msec) working memory load [$t(31) = 0.74$, $p = .46$]. This finding remains even when performance on the high-memory-load secondary task is included as a covariate in an ANCOVA, showing that the effect of memory load on flanker task performance is not significant [$F(1,31) = 0.55$, $p = .46$; covariate, $F(1,30) =$ 0.25, $p = .62$] and when calculated as a proportion of mean RT [low, 20.6%; high, 21.9%; $t(31) = 0.63$, $p = .53$; $F(1,31) = 0.40$, $p = .53$; covariate, $F(1,30) = 0.24$, $p =$ .63], suggesting that individual variation in the ability to perform the memory load task did not affect performance on the flanker task.

Discussion

The results of the first experiment provide qualified support for the application of load theory to audition—at least, in a spatial word recognition task. The effect of increasing perceptual load is unequivocal: In Experiment 1, increased perceptual load resulted in decreased interference from competing speech, precisely as predicted by load theory and in apparent contradiction to the findings of Gomes et al. (2008). In contrast, not only is the lack of an effect of increasing WM load on interference from competing speech contrary to predictions of load theory, it also does not seem commensurate with expectations based on other studies that have clearly demonstrated effects of WM load in other spoken word recognition tasks (e.g., Francis, Nusbaum, & Fenn, 2007; Nusbaum & Morin, 1992).

There are at least two reasons that could plausibly explain the finding of perceptual load effects in the present study, but not in that of Gomes et al. (2008). First, the task used by Gomes and colleagues may have manipulated load on a mechanism for object formation rather than object selection, and this mechanism may not be as subject to load effects. Following Shinn-Cunningham (2008), auditory selective attention involves at least two processes: object formation (assigning distinct perceptual events to the same real-world source) and object selection (prioritizing one object over another for further processing). For example, in order to selectively attend to the speech of one talker over another, listeners must first process the auditory scene in order to group acoustic components (e.g., harmonics) of each voice appropriately (cf. Bregman, 1990). Only after distinct perceptual representations of each voice have been formed are these "objects" available to the influences of selective attention (although see Griffiths & Warren, 2004, for a discussion of the difficulty of defining auditory objects). If the task used by Gomes et al. primarily depended on listeners' ability to segregate the auditory scene, imposing a greater perceptual load on this mechanism may not have affected the participants' ability to selectively attend to distinct auditory objects in the same way as would a task that primarily loaded object selection. In support of this interpretation, in the Gomes et al. study, load was increased by decreasing the interval between individual tones in each stream. Following the theory of auditory scene analysis (Bregman, 1990), greater proximity in time can actually improve listeners' ability to "stream" individual tones into a single auditory object; in other words, although the decreased ISI may have increased the load on mechanisms for making decisions about individual tones in a stream (as intended by Gomes et al., 2008), it may also have *decreased* processing demand for systems of organizing the auditory scene by facilitating the grouping of high tones into one object and low tones into another. Finally, in light of the findings of Alain and Izenberg (2003) suggesting that segregation of simultaneous objects (as in the present experiment) may be affected by capacity limitations less than is the segregation of objects defined according to temporal properties (as in the Gomes et al., 2008, study), further research is clearly necessary to determine whether, or to what degree, processes of object formation are affected by cognitive load.

On the other hand, it is also possible that the present finding of decreased interference with increased perceptual load is a function of the fundamentally spatial nature of the present task, unrelated to the auditory modality in which it was presented. Gomes et al. (2008) used stimuli that were differentiable as distinct auditory objects only on the basis of frequency, not space. It is possible that the perceptual load effect predicted by load theory arises only in conditions in which target and distractor objects are spatially differentiable. Although Gomes et al. argued that the findings of Barnhardt, Ritter, and Gomes (2008) suggest otherwise, at least in the visual modality, the next experiment was designed to address this question directly.

Interestingly, the spatial nature of the present task could also explain the *failure* to find the expected effect of *memory* load in the present experiment. If spatial separation of the target and distractor resulted in an improvement in the efficiency of auditory object formation (as was suggested, e.g., by Freyman et al., 1999), this in turn may have reduced overall demand on cognitive processing—either directly, by reducing early demand on memory systems and thereby allowing them to be applied more effectively to higher level processing (i.e., distractor inhibition), or indirectly, by providing more distinct representations of auditory objects and thereby facilitating the subsequent distribution of selective attention among them. In either case, spatial separation may have reduced the cognitive capacity demand of the primary flanker task to such a degree that the manipulation of WM capacity via the secondary memory load task was unable to show a significant effect. Thus, by presenting target and distractor stimuli from the same perceived location in both perceptual and working memory load conditions, the next experiment was able to test both predictions relevant to reconciling the results of the first experiment with those of Gomes et al. (2008) and to examine a possible explanation for the failure to find an effect of WM load in Experiment 1.

EXPERIMENT 2

The second experiment was designed to evaluate hypotheses related to the role of spatial separation of target and flanker derived from the results of Experiment 1. If the perceptual load effects observed in the first experiment disappear when all stimuli are presented from the same perceived spatial location, this would support the hypothesis that these effects resulted from the operation of a mechanism unique to spatially directed attention, thereby reconciling the present results with those of Gomes et al. (2008). Conversely, if, by presenting all stimuli from the same perceived spatial location, memory load effects are observed (in contrast to Experiment 1), it would support the hypothesis that spatial release from informational masking played a role in obscuring the effects of memory load on speech perception in the first experiment. Therefore, the second experiment was designed as an exact replication of the first, with two exceptions. First, all the stimuli were presented diotically, meaning that all were perceived as originating at 0º (straight ahead, or in the center of the listener's head). Second, because it was no longer possible to designate the target according to spatial location, the listeners were instructed instead to identify the speech of either the male or the female talker (counterbalanced across listeners).

Method

Participants

Forty-five participants were recruited for this experiment ($N = 45$; 23 women, 22 men; mean age = 21 years, 4 months). All met the same criteria as the individuals included in Experiment 1, except that 6 were left-handed. None of the participants in Experiment 2 had participated in Experiment 1. All testing was conducted under a protocol approved by the Committee for the Protection of Human Research Subjects at Purdue University, and all the participants were paid for their time.

Stimuli

All the stimuli were identical to those in Experiment 1, except that target and flanker (distractor) wave files were combined at equal intensity into both channels of the stereo .wav file.

Procedure

The procedure was identical to that described in Experiment 1, except that the listeners were instructed to attend to either the male or the female voice as the target. Approximately equal numbers of male and female participants attended to each voice (female participants, female target = 12, male target = 11; male participants, female target = 12, male target = 10).

Results

RTs and interference scores for both the perceptual and memory load tasks are shown in Figure 4. As in Experiment 1, the results from the two tasks were analyzed separately.

Perceptual Load

Secondary task performance was comparable to that shown in Experiment 1, with participants averaging a d' of 2.39 in the congruent condition (2.88 under low load, 1.90 in the high) and 2.13 in the incongruent condition (2.65 in the low-load condition, 1.61 in the high-load condition). A two-way repeated measures ANOVA showed significant effects of load level [$F(1,44) = 51.49$, $p < .001$], but not of trial type [$F(1,44) = 3.35$, $p = .07$], and no interaction [$F(1,44) = 0.04$, $p = .84$]. Although the participants did have more difficulty with the high- than with the low-load task, primary task properties did not affect performance on the secondary task.

On the primary task, the participants were again accurate at recognizing the target word, although somewhat less so than in the first experiment, averaging 87.8% correct in the congruent condition (90.9% in the low-load blocks, 84.7% in the high) and 81.0% correct in the incongruent condition (85.3% in the low-load blocks, 76.7% in the high). A two-way repeated measures ANOVA showed significant effects of trial type [$F(1,44) = 12.49$, $p = .001$] and of load level [$F(1,44) = 14.79$, $p < .001$] but no interaction [$F(1,44) = 0.39$, $p = .54$]. Although the participants became slightly less accurate at recognizing the target word under high load and on incongruent trials, the effect of load did not differ between congruent and incongruent trials at either high or low load levels.

With respect to RTs, listeners were again faster on congruent trials ($M = 1{,}222$ msec; 1,099 msec on low-load

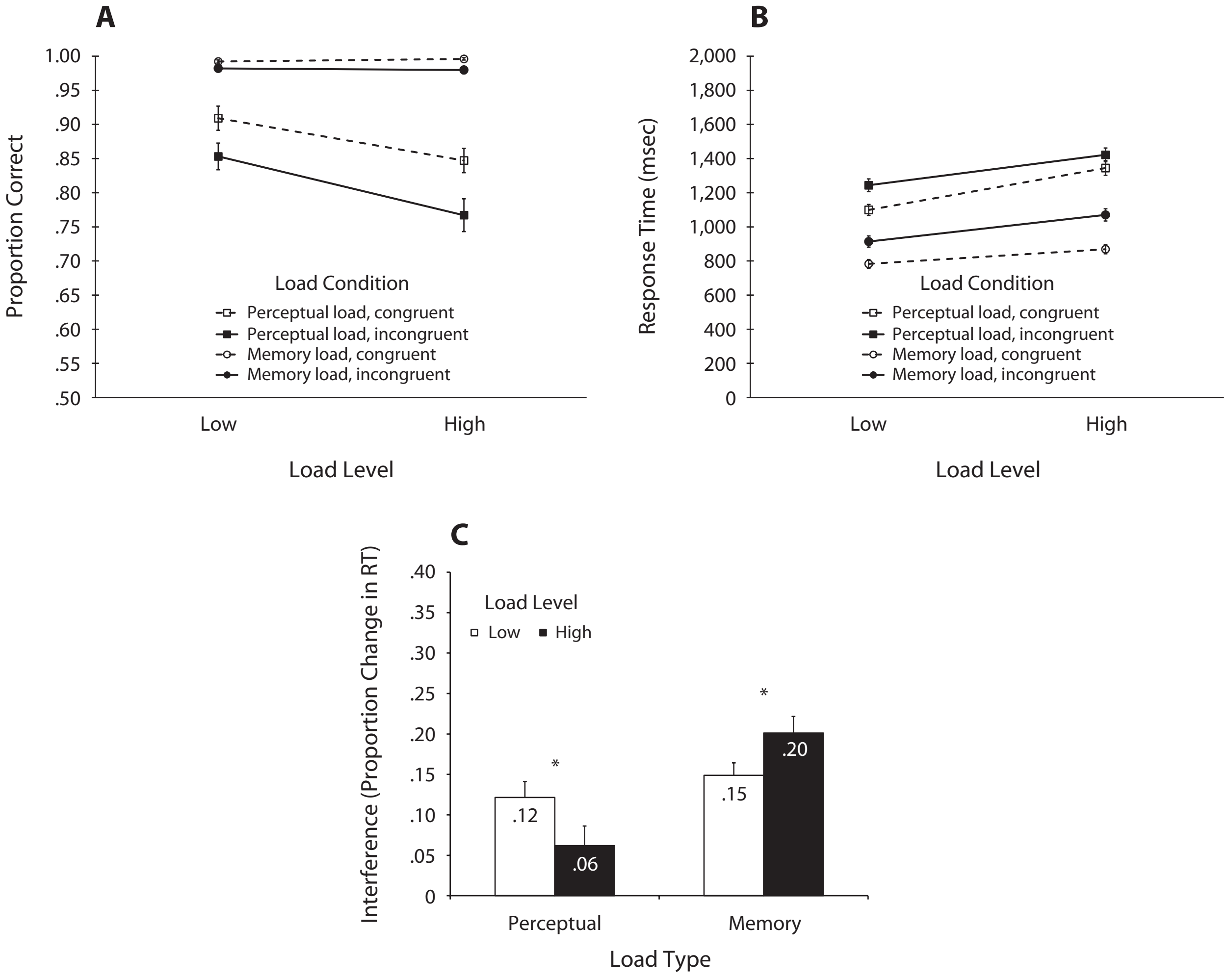

Figure 4. Results for Experiment 2, showing proportions of correct responses (A), response times (RTs) on correct responses for both congruent and incongruent trials of the auditory flanker task under low and high perceptual and memory load (B), and interference (proportion of increase in RT from congruent to incongruent conditions) under low and high perceptual and memory load (C). Error bars indicate 1 *SE*.

trials, 1,345 msec on high) than on incongruent ones (M = 1,333 msec; 1,244 msec on low-load blocks, 1,422 msec on high), and a two-way repeated measures ANOVA showed significant main effects of trial type [$F(1,44) = 8.29$, $p = .006$] and load level [$F(1,44) = 30.50$, $p < .001$] and no significant interaction between the two [$F(1,44) = 0.78$, $p = .38$]. As in Experiment 1, interference was quantified as the difference between incongruent and congruent RTs on correct responses, and a Student's t test of dependent samples was conducted to compare interference in the high-load condition (144 msec) with that in the low (77 msec). The results were nearly significant [$t(44) = 1.92$, $p = .06$]. Analyzing the same data in terms of proportion of mean RT for the condition showed a significant effect between low (12.1%) and high (6.2%) loads [$t(44) = 2.31$, $p = .03$]. Since, as was discussed previously, the normalized results are probably more reliable (and the absolute difference between raw values in milliseconds is actually greater in this experiment than in the first one), these results again support the predictions of load theory: Increased perceptual load results in a decrease in interference in a speech perception task. These findings are comparable or even slightly improved when secondary task performance, in terms of d' under high load, is included as a covariate in an ANCOVA [$F(1,44) = 3.80$, $p = .054$; covariate, $F(1,43) = 1.25$, $p = .27$] [for percentages, $F(1,44) = 5.34$, $p = .03$; covariate, $F(1,43) = 1.68$, $p = .20$], suggesting that individual variation in the ability to perform the high-perceptual-load secondary task may have affected the listeners' ability to perform the primary task, but only slightly.

Working Memory Load

The listeners were able to recall digits nearly as well as in Experiment 1, both on congruent trials (95.3% in the low-load blocks, 89.3% in the high) and on incongruent ones (93.7% in the low-load blocks, 87.5% in the high). The results of a two-way repeated measures ANOVA showed a significant effect of load level [$F(1,44) = 32.25$, $p < .001$] but no effect of trial type [$F(1,44) = 2.42$, $p = .13$] and no interaction between the two [$F(1,44) = 0.01$, $p = .91$].

The listeners in Experiment 2 were again quite successful at recognizing target words. Response accuracy was 99.4% correct on congruent trials (99.2% correct in low-load blocks, 99.6% in high) and 98.1% correct on incongruent ones (98.2% in low-load blocks, 98.0% in high). A two-way repeated measures ANOVA of these scores showed a significant effect of trial type [$F(1,44) = 14.19$, $p < .001$], but not of load [$F(1,44) = 0.04$, $p = .84$], and no interaction [$F(1,44) < 0.80$, $p = .38$].

With respect to RT, the listeners were again faster on congruent trials (M = 826 msec; low = 783 msec, high = 869 msec) than on incongruent ones (M = 992 msec; low = 914 msec, high = 1,070 msec), and a two-way repeated measures ANOVA with two levels of trial type and two of load level showed a significant main effect of trial type [$F(1,44) = 29.84$, $p < .001$] and of load [$F(1,44) = 15.78$, $p < .001$; low = 849 msec, high = 970 msec], but no interaction [$F(1,44) = 1.33$, $p = .25$]. Again, a paired-comparison Student's t test was conducted on interference scores. In this experiment, the results showed a significant difference between interference under low (131 msec), as compared with high (202 msec), WM load [$t(44) = 4.28$, $p < .001$]. This finding remains significant even when performance on the high-memory-load secondary task is included as a covariate in an ANCOVA [$F(1,44) = 18.34$, $p < .001$; covariate $F(1,43) = 2.81$, $p = .10$] and when calculated as a percentage of mean RT for the condition [low = 14.9%, high = 20.1%; $t(44) = 3.15$, $p = .003$; ANCOVA, $F(1,44) = 9.89$, $p = .003$; covariate $F(1,43) = 3.33$, $p = .08$], suggesting that individual variation in the ability to perform the memory load task did not affect performance on the primary task but that, when both target and distractor are perceived as coming from the same location, memory load does have a significant effect on processing interference, as is predicted by load theory.

Comparison of RTs With Experiment 1

One implication of the differences in the effects of WM load between Experiments 1 and 2 is that degree of perceived spatial separation of the target and distractor may affect the WM demands of processing the speech of one talker while ignoring that of another. To examine this possibility more directly, a mixed factorial ANOVA was conducted for RT with one between-groups factor with two levels (spatial distance) and two repeated measures, each with two levels (trial type, congruent vs. incongruent; load level, low vs. high). The results showed significant main effects of trial type [$F(1,75) = 64.32$, $p < .001$] and of load level [$F(1,75) = 9.32$, $p = .003$] and a significant interaction between group and load level [$F(1,75) = 4.23$, $p = .04$]. The group × load interaction is shown in Figure 5. No other main effects or interactions were significant at the .05 level. A post hoc (Tukey HSD) analysis of the group × load interaction showed a marginally significant difference ($p = .06$) in RTs between the one-

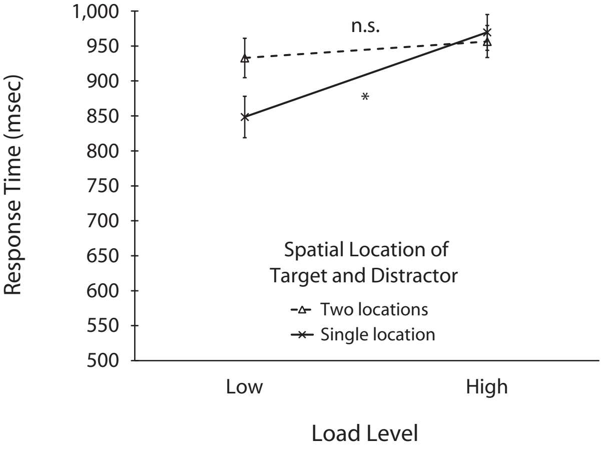

Figure 5. Comparison of response times on correct flanker trials under low and high working memory load for the participants in Experiment 1 (two locations) and Experiment 2 (single location). Error bars indicate 1 *SE*.

and two-location groups under low memory load (849 and 933 msec, respectively), but not under high (970 and 957 msec, respectively). Similarly, increased memory load induced a significant ($p < .001$) increase in RT in the one-location condition (from 849 to 970 msec), but not in the two-location condition (from 933 to 957 msec).[3]

Discussion

The finding that perceptual load effects on processing of speech in competing speech are maintained when all the stimuli are presented from the same perceived location suggests that the apparent differences between the conclusions of Gomes et al. (2008) and those of the present study did not result from the spatial separation of the target and masker in Experiment 1. In addition, the appearance of memory load effects when all the stimuli are presented from the same perceived location supports the hypothesis that spatial release from informational masking may result, at least in part, from a reduction of WM demand for directing attention to targets and away from distractors perceived as coming from separate locations, as opposed to a single one.

GENERAL DISCUSSION

The pattern of results exhibited in the two experiments presented here provides support for the application of load theory to the study of speech perception and provides new insight into the nature of the mechanisms involved in auditory selective attention, both at a perceptual and at a cognitive level.

Perceptual Load

Gomes et al. (2008) argued that their failure to find an effect of perceptual load in nonspeech auditory perception reflects fundamental differences in the ecological demands on the auditory and visual systems, so that the auditory system may not automatically allocate all available resources to objects in the auditory scene in the same way that the visual system does. The present findings do not directly contradict this hypothesis: It is still possible that, at very early stages of auditory scene analysis, auditory attentional capacity need not be fully allocated, such that, with the relatively simple auditory scene presented by Gomes et al., no effects of increasing perceptual load are found. Electrophysiological and behavioral studies of auditory scene analysis similarly suggest that early segregation of auditory objects occurs relatively independently of attention (cf. Alain & Izenberg, 2003, and the discussion by Alain, 2007). For example, increasing attentional load does not affect listeners' ability to identify the presence of a mistuned harmonic (Alain & Izenberg, 2003), but it does affect identification of two simultaneously presented vowels (Alain et al., 2005). Both of these tasks (perceiving mistuned harmonics and identification of concurrent vowels) depend on successfully analyzing the auditory scene, possibly by adjusting a harmonic template to fit the individual components of the complex wave in order to determine whether the complex wave likely contains sounds from one or more sources (Alain et al., 2005). The crucial difference between the two tasks is that identifying concurrent vowels also involves identification and categorization of the two vowels, whereas recognizing a mistuned harmonic does not. Following the terminology of Shinn-Cunningham (2008), the first process (adjusting a harmonic template) is a fundamental component of object formation, whereas the second (vowel categorization) also requires object selection. Thus, the work of Alain and colleagues suggests that the process of object formation may not be affected by attentional load, whereas object selection is. On the basis of this reasoning, it seems likely that the task used by Gomes et al. may have depended to a greater degree on object formation than on object selection, making it relatively impervious to manipulations of attentional load, exactly as reported. In contrast, the auditory flanker task used in the present experiments depends quite heavily on processes of object selection (in addition to those of object formation) and, therefore, should be expected to exhibit a greater dependence on the availability of attentional resources. Thus, the present results are strongly consistent with the assumption of load theory—namely, that perceptual-processing capacity must be fully allocated to the scene. However, they raise the possibility that this dependence is manifest only once the objects in the scene are well specified by processes of object formation.

Cognitive Load

Differences between the results of the present first and second experiments also suggest that spatial release from informational masking may involve, at least in part, a reduction in the cognitive demands incurred by selectively attending to one object and ignoring the other when the two objects are perceived as arising from distinct spatial locations. That is, presenting spatially separated targets and distractors (in Experiment 1) seems to have reduced demand on cognitive capacity to a degree sufficient to eliminate interaction with the secondary WM load task. The fact that a trend in the expected direction was still observed in Experiment 1 suggests that this reduction in demand may be a matter of degree, not an absolute change in processing strategies or mechanisms, and further suggests that it might be possible to induce an effect of WM load, even with spatially separated targets and distractors, as long as the secondary task WM load is sufficiently high.

There are two possible explanations for such facilitation. On the one hand, it is possible that spatial separation works primarily on processes of object formation, improving initial analysis and/or separation of frequency components of the two signals, allowing for more distinct representations of each object (i.e., involving more focused, less overlapping patterns of neural activity corresponding to each object), and thereby supporting more effective use of limited WM resources required to manipulate them during subsequent processing (i.e., for categorization/identification).[4] On the other hand, it is also possible that spatial separation reduces informational masking not by improving the initial formation of distinct objects per se, but rather by augmenting the representation of objects with additional cues (i.e., spatial location). Richer object representations could facilitate object selection directly by providing more, and more distinctive, cues on the basis of which each object can be more easily selected, or inhibited, or recalled from memory (Nairne, 2003; Neath, 2000).

Following Ihlefeld and Shinn-Cunningham (2008), spatial separation of targets and distractors can improve both the process of object formation and that of object selection under certain conditions. When listeners are directed to selectively attend to a spatially defined target (as in Experiment 1) or to a target defined both spatially and in terms of timbre (as in Experiment 2), both object formation and object selection are facilitated. However, when listeners are directed to attend to a target defined *only* in terms of timbre (and at unpredictable locations), separation of target and masker improves only object formation (not selection). Thus, because target and distractor locations were constant across trials in both experiments presented here, the listeners had implicit (Experiment 2) or explicit (Experiment 1) knowledge of target location (Experiment 1) or location and timbre (Experiment 2), and therefore, conditions were always sufficient to allow spatial separation to facilitate both object formation and selection. Further research will be necessary to determine whether the benefit to WM demand provided by spatial separation of targets and distractors results from an interaction with processes of auditory object formation, selection, or both.

Future Research

Although the precise mechanism has not yet been identified, the results of the present study suggest that one way in which spatial separation of targets and distracting speech may facilitate speech recognition in the presence of competing speech is by reducing the WM demand for processing target speech and/or inhibiting the processing of irrelevant speech. This finding may be particularly relevant for future research on factors affecting older listeners' ability to recognize speech in competing speech.

Previous research has provided mixed results regarding the role of cognitive factors in explaining older adults' difficulties with understanding speech in adverse conditions, such as in the presence of noise, reverberation, or competing speech, with the general consensus being that auditory acuity plays the strongest role and age-related cognitive changes may or may not be relevant (cf. Akeroyd, 2008, for a review). However, those studies that show a relationship between independent objective measures of cognition and perception of speech in noise or competing speech typically have included speech perception tasks that depend heavily on WM. For example, Pichora-Fuller et al. (1995) used a combined sentence verification and word recall task and showed that older listeners hearing sentences in babble noise remembered fewer items *that they had perceived correctly*, as compared with younger listeners. On the basis of these findings, they argued that older adults have increased difficulty with understanding speech in adverse conditions because they reallocate higher level cognitive resources to compensate for age-related loss of low-level auditory acuity (see also McCoy et al., 2005; Rabbitt, 1991; Tun et al., 2002; Wingfield, Tun, & McCoy, 2005). Similarly, Humes et al. (2006) found much stronger effects of age on a speech-in-speech task involving divided attention (more demanding) than one involving selective attention (less demanding). Moreover, individual differences in speech perception performance were correlated with WM capacity as measured by a digit span task, both under divided attention conditions (high WM demand) and in selective attention conditions with high variability in masking talker identity (with greater uncertainty again leading to higher cognitive demand). Finally, Murphy, Daneman, and Schneider (2006) observed that older adults were worse than younger adults at remembering details about the content of spoken dialogues in noise when the two talkers were separated in space, but this age-related advantage disappeared when the two talkers were presented from the same location. Murphy et al. argued that older adults' memories for dialogue details in the 45º and 0º separation conditions were equivalent because they were unable to make use of spatial cues, and this is supported by the observation that younger adults' performance fell to be equivalent to that of older adults when spatial cues were eliminated in the 0º condition. Given that the outcome measure in the studies discussed here depended fundamentally on memory of a spoken message, these findings support the hypothesis that elimination of spatial cues to talker segregation (whether due to age or by intentionally colocating talkers) increases WM demand for online speech processing and, thereby, reduces the availability of WM capacity for message comprehension and memory. However, the presence or absence of spatial separation seems to have little effect on the interaction between perceptual load and interference from a distracting talker, suggesting that the systems underlying perceptual and cognitive capacity are not isomorphic.

CONCLUSIONS

In a dual-task paradigm, increasing perceptual and cognitive (WM) load had different effects on the degree to which the speech of an irrelevant talker interfered with processing of a target talker's speech, in a manner consistent with the predictions of load theory (Lavie, 2005). As perceptual load increased, irrelevant speech interfered less with target word recognition. In contrast, as cognitive load increased, listeners' ability to filter out distracting speech decreased. The effects of increasing cognitive load appear to be modulated by spatial separation between the target and distractor: The presence of spatial separation may reduce the overall WM demands of selecting one talker over another, resulting in a reduction, but not necessarily elimination, of the effect of increased cognitive load on distractor interference. Further research is necessary to investigate the mechanisms by which spatial separation affects WM demands for processing speech in competing speech and to investigate the degree to which limited perceptual and cognitive capacities may be related to one another in speech perception.

AUTHOR NOTE

This research was funded by Grant R03DC006811 from NIH/NIDCD. The author thanks Michelle Morgan, Amber Page, Erica Hartland, and Katie Connell for data collection and Hilary Gomes, Howard Nusbaum, and two anonymous reviewers for helpful comments on earlier drafts of the manuscript. Correspondence concerning this article should be addressed to A. L. Francis, Speech, Language, and Hearing Sciences, Purdue University, 500 Oval Drive, West Lafayette, IN 47907 (e-mail: francisa@purdue.edu).

REFERENCES

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology*, **47**(Suppl. 2), S53-S71.

Alain, C. (2007). Breaking the wave: Effects of attention and learning on concurrent sound perception. *Hearing Research*, **229**, 225-236.

Alain, C., & Izenberg, A. (2003). Effects of attentional load on auditory scene analysis. *Journal of Cognitive Neuroscience*, **15**, 1063-1073.

Alain, C., Reinke, K., He, Y., Wang, C., & Lobaugh, N. (2005). Hearing two things at once: Neurophysiological indices of speech segregation and identification. *Journal of Cognitive Neuroscience*, **17**, 811-818.

Allport, A. (1993). Attention and control: Have we been asking the wrong question? A critical review of twenty-five years. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 183-210). Cambridge, MA: MIT Press.

Barnhardt, J., Ritter, W., & Gomes, H. (2008). Perceptual load affects spatial and nonspatial visual selection processes: An event-related brain potential study. *Neuropsychologia*, **46**, 2071-2078.

Boersma, P., & Weenink, D. (2007). Praat: Doing phonetics by computer (Version 4.6.08) [Computer program]. Available at www.praat.org.

Brand-D'Abrescia, M., & Lavie, N. (2008). Task coordination between and within sensory modalities: Effects on distraction. *Perception & Psychophysics*, **70**, 508-515.

Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.

Carpenter, P. A., Just, M. A., Keller, T. A., Eddy, W., & Thulborn, K. (1999). Graded functional activation in the visuospatial system with the amount of task demand. *Journal of Cognitive Neuroscience*, **11**, 9-24.

Chan, J. S., Merrifield, K., & Spence, C. (2005). Auditory spatial attention assessed in a flanker interference task. *Acta Acustica United With Acustica*, **91**, 554-563.

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

de Fockert, J. W., Rees, G., Frith, C. D., & Lavie, N. (2001). The role of working memory in visual selective attention. *Science*, **291**, 1803-1806.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, **70**, 80-90.

Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature*, **381**, 66-68.

Driver, J. (2001). A selective review of selective attention research from the past century. *British Journal of Psychology*, **92**, 53-78.

Francis, A. L., & Nusbaum, H. C. (2009). Effects of intelligibility on working memory demand for speech perception. *Attention, Perception, & Psychophysics*, **71**, 1360-1374.

Francis, A. L., Nusbaum, H. C., & Fenn, K. (2007). Effects of training on the acoustic–phonetic representation of synthetic speech. *Journal of Speech, Language, & Hearing Research*, **50**, 1445-1465.

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., 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., Mason, C. R., & Kidd, G., Jr. (2007). Task-dependent costs in processing two simultaneous auditory stimuli. *Perception & Psychophysics*, **69**, 757-771.

Gomes, H., Barrett, S., Duff, M., Barnhardt, J., & Ritter, W. (2008). The effects of interstimulus interval on event-related indices of attention: An auditory selective attention test of perceptual load theory. *Clinical Neurophysiology*, **119**, 542-555.

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience*, **5**, 887-892.

Humes, L. E. (2002). Factors underlying the speech-recognition performance of elderly hearing-aid wearers. *Journal of the Acoustical Society of America*, **112**, 1112-1132.

Humes, L. E., Lee, J. H., & Coughlin, M. P. (2006). Auditory measures of selective and divided attention in young and older adults using single-talker competition. *Journal of the Acoustical Society of America*, **120**, 2926-2937.

Ihlefeld, A., & Shinn-Cunningham, B. (2008). Disentangling the effects of spatial cues on selection and formation of auditory objects. *Journal of the Acoustical Society of America*, **124**, 2224-2235.

Keller, T. A., Carpenter, P. A., & Just, M. A. (2001). The neural bases of sentence comprehension: A fMRI examination of syntactic and lexical processing. *Cerebral Cortex*, **11**, 223-237.

Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2007). Informational masking. In W. Yost (Ed.), *Springer handbook of auditory research: Vol. 29. Auditory perception of sound sources* (pp. 143-190). New York: Springer.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 451-468.

Lavie, N. (2000). Selective attention and cognitive control: Dissociating attentional functions through different types of load. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 175-194). Cambridge, MA: MIT Press.

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.

Madden, D. J. (2001). Speed and timing of behavioral processes. In J. E. Birren & K. W. Shaie (Eds.), *Handbook of the psychology of aging* (5th ed., pp. 288-312). San Diego: Academic Press.

McCoy, S. I., Tun, P. A., Cox, I. C., Colangelo, M., Steward, R. A., & Wingfield, A. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *Quarterly Journal of Experimental Psychology*, **58A**, 22-33.

Moore, B. C. J. (2003). *An introduction to the psychology of hearing* (5th ed.). Amsterdam: Academic Press.

Murphy, D. R., Daneman, M., & Schneider, B. A. (2006). Why do older adults have difficulty following conversations? *Psychology & Aging*, **21**, 49-61.

Nairne, J. S. (2003). Sensory and working memory. In A. F. Healy & R. W. Proctor (Eds.), *Comprehensive handbook of psychology: Vol. 4. Experimental psychology* (pp. 423-444). New York: Wiley.

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

Nusbaum, H. C., & Magnuson, J. (1997). Talker normalization: Phonetic constancy as a cognitive process. In K. Johnson & J. W. Mullenix (Eds.), *Talker variability in speech processing* (pp. 109-132). San Diego: Academic Press.

Nusbaum, H. C., & Morin, T. M. (1992). Paying attention to differences among talkers. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaki (Eds.), *Speech perception, production, and linguistic structure* (pp. 113-134). Tokyo: Omsha.

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America*, **97**, 593-608.

Rabbitt, P. M. (1968). Channel capacity, intelligibility and immediate memory. *Quarterly Journal of Experimental Psychology*, **20**, 241-248.

Rabbitt, P. M. (1991). Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. *Acta Oto-Laryngologica Supplement*, **476**, 167-176.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, **114**, 510-532.

Rees, G., Frith, C., & Lavie, N. (1997). Modulating irrelevant motion perception by varying attentional load in an unrelated task. *Science*, **278**, 1616-1619.

Rees, G., Frith, C., & Lavie, N. (2001). Processing of irrelevant visual motion during performance of an auditory attention task. *Neuropsychologia*, **39**, 937-949.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime 1.2 [Computer software]. Pittsburgh: Psychology Software Tools.

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, **12**, 182-186.

Treisman, A. M. (1964). The effect of irrelevant material on the efficiency of selective listening. *American Journal of Psychology*, **77**, 533-546.

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

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, **12**, 97-136.

Treisman, A. [M.], & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 459-478.

Tun, P. A., O'Kane, G., & Wingfield, A. (2002). Distraction by competing speech in younger and older listeners. *Psychology & Aging*, **17**, 453-467.

Vongpaisal, T., & Pichora-Fuller, M. L. (2007). Effect of age on F_0 difference limen and concurrent vowel identification. *Journal of Speech, Language, & Hearing Research*, **50**, 1139-1156.

Wingfield, A., & Grossman, M. (2006). Language and the aging brain: Patterns of neural compensation revealed by functional brain imaging. *Journal of Neurophysiology*, **96**, 2830-2839.

Wingfield, A., Tun, P. A., & McCoy, S. L. (2005). Hearing loss in older adulthood: What is it and how it interacts with cognitive performance. *Current Directions in Psychological Science*, **14**, 144-148.

Wong, P. C. M., Nusbaum, H. C., & Small, S. L. (2004). Neural bases of talker normalization. *Journal of Cognitive Neuroscience*, **16**, 1173-1184.

Yost, B. (2006, March). *Informational masking: What is it?* Paper presented at the 2006 Computational and Systems Neuroscience (Cosyne) Meeting. Available for download at www.isr.umd.edu/Labs/NSL/Cosyne/Yost.htm.

NOTES

1. Because of the amplitude modulation, the AM signals did not reach peak amplitude until 71 msec into the tone signal.

2. It is important to note that this manipulation also probably imposed a greater WM load in the high- than in the low-perceptual-load condition, since the high-load condition required listeners to retain a more complex rule in memory. Although the written presentation of the response rule should have helped alleviate this burden, the confounding of WM load and perceptual load in this task certainly cannot be ruled out as a factor, either in the present results or in those of other, similarly structured visual tasks (e.g., Lavie, 1995).

3. Note that log transformation of RTs, advisable when effect probabilities may be marginal (Ratcliff, 1993), results in a maintenance or even strengthening of all the results, including the group effect, which becomes close to marginally significant [group, $F(1,75) = 2.78$, $p = .1$; trial type, $F(1,75) = 62.06$, $p < .001$; load level, $F(1,75) = 9.85$, $p = .002$; group × load, $F(1,75) = 4.82$, $p = .03$]. Relevant pairwise comparisons likewise remain or become significant (single location vs. two locations at low load, $p = .04$; low vs. high load with stimuli at a single location, $p = .001$). As with the untransformed data, no other main effects, interactions, or relevant pairwise comparisons were significant at the $\leq .05$ level.

4. In this respect, it is important to note that the target may also have been perceived as louder than the distractor in the first, but not the second, experiment. In Experiment 1, MLD effects may have increased the perceived loudness of the diotically presented stimulus vis-a-vis the monaural one, whereas in Experiment 2, the target and distractor were equally loud in both ears. Although an attempt was made to reduce these effects, further research is necessary to distinguish between effects of intensity versus spatial distinctiveness on WM demand for selectively attending to speech in competing speech.

(Manuscript received June 30, 2008;
revision accepted for publication October 15, 2009.)