

Contextual effects in the identification of nonspeech auditory patterns

Gerald Kidd Jr.^{a)}

*Department of Speech, Language and Hearing Sciences and Hearing Research Center, Boston University,
635 Commonwealth Avenue, Boston, Massachusetts 02215*

Virginia M. Richards

Department of Cognitive Sciences, University of California, Irvine, California 92697

Timothy Streeter and Christine R. Mason

*Department of Speech, Language and Hearing Sciences and Hearing Research Center, Boston University,
635 Commonwealth Avenue, Boston, Massachusetts 02215*

Rong Huang

Department of Cognitive Sciences, University of California, Irvine, California 92697

(Received 6 February 2011; revised 5 October 2011; accepted 7 October 2011)

This study investigated the benefit of *a priori* cues in a masked nonspeech pattern identification experiment. Targets were narrowband sequences of tone bursts forming six easily identifiable frequency patterns selected randomly on each trial. The frequency band containing the target was randomized. Maskers were also narrowband sequences of tone bursts chosen randomly on every trial. Targets and maskers were presented monaurally in mutually exclusive frequency bands, producing large amounts of informational masking. Cuing the masker produced a significant improvement in performance, while holding the target frequency band constant provided no benefit. The cue providing the greatest benefit was a copy of the masker presented ipsilaterally before the target-plus-masker. The masker cue presented contralaterally, and a notched-noise cue produced smaller benefits. One possible mechanism underlying these findings is auditory “enhancement” in which the neural response to the target is increased relative to the masker by differential prior stimulation of the target and masker frequency regions. A second possible mechanism provides a benefit to performance by comparing the spectrotemporal correspondence of the cue and target-plus-masker and is effective for either ipsilateral or contralateral cue presentation. These effects improve identification performance by emphasizing spectral contrasts in sequences or streams of sounds. © 2011 Acoustical Society of America. [DOI: 10.1121/1.3658442]

PACS number(s): 43.66.Dc, 43.66.Lj [LD]

Pages: 3926–3938

I. INTRODUCTION

The success that a listener achieves in processing a complex sound, such as hearing out an element or feature of the sound or extracting enough information about the sound to identify it, can be influenced by other surrounding sounds (we will refer to the general influence of sounds surrounding a “target” sound as “contextual effects”). For example, when the task of the listener is to detect the presence of a tone embedded in a masker, *cuing* the target tone may, under certain circumstances, make it easier to hear and improve target-to-masker ratio (T/M) at threshold significantly (e.g., Richards and Neff, 2004). A similar finding has been reported for more complex tasks such as speech recognition. Again, under appropriate conditions, a cue or “precursor” to (e.g., Holt *et al.*, 2000) or priming of (e.g., Freyman *et al.*, 2004; Sheldon *et al.*, 2008) a speech target may influence how the sound is perceived and often provides a significant aid to speech identification performance. In certain cases, such as in highly

uncertain listening conditions dominated by informational masking (e.g., Kidd *et al.*, 2008a), cuing the masker may be more beneficial than cuing the target (e.g., Richards and Neff, 2004; Richards *et al.*, 2004). Furthermore, contextual effects may be detrimental to performance as in the case of, for example, forward masking (see Carlyon, 1989, for work related to forward masking and enhancement; for general examples, see Munson and Gardner, 1950; Kidd and Feth, 1981; Plack and Oxenham, 1998).

Although contextual effects may provide a significant benefit to performance in a wide variety of experiments, many of these effects currently are not well understood, and questions remain about the extent to which they arise from common underlying mechanism(s). One important issue concerns how the benefits of context depend on the nature of the task and the type of stimulus that is employed in the task. The present study examines contextual effects in one specific instance in which the ability to identify the elements of a set of previously learned nonspeech patterns is influenced by a preceding cue. This instance is of interest because the task is intermediate to detection and speech recognition providing a link between at-threshold and suprathreshold processes

^{a)}Author to whom correspondence should be addressed. Electronic mail: gkidd@bu.edu

without the complication of the linguistic factors governing speech perception. Moreover, for reasons explained in the following text, the benefit of context due to *a priori* cuing is examined under masked conditions having a high degree of stimulus uncertainty, producing masking that is predominantly informational in nature. The broad conclusion is that there are at least two mechanisms underlying the beneficial effects of cuing found under these conditions. The generalization of these mechanisms to other examples among the wide range of contextual effects is not fully clear but is considered further in the following text.

It is beyond the scope of this article to review the broad topic of contextual effects in sequences of sounds. The focus here is on a subset of studies and issues that are most closely related to the auditory “enhancement effect.” The original work on the enhancement effect described advantages for the task of detection of narrow-band nonspeech targets embedded in simultaneous “maskers” that were either nonspeech or speech-like (harmonic complexes) sounds (Viemeister, 1980). The “enhancement” referred to a reduction in masked threshold for the target due to the presence of a precursor sound or cue.¹ A subjective correlate to these psychophysical unmasking demonstrations had been reported earlier by Zwicker (1964) in which the enhanced tone (parallel to target in masking experiments) appeared to segregate from the complex (masker) in which it was embedded. Zwicker observed that the perception of the enhanced tone persisted over several seconds, an observation with which Viemeister (1980) concurred. There is a pervasive issue, though, about the extent to which the persistence of the subjective impression of an enhanced tone means that its masked detectability, or the effectiveness of the enhanced tone when used as a masker, persists as well, which is considered again later in Sec. IV. Viemeister (1980) described a series of experiments in which the prior presentation of a stimulus with a spectral gap—either a notched-filtered noise or a harmonic complex with one component excised—lowered detection thresholds in the frequency region corresponding to the gap when the full stimulus (target plus masker) followed. In his study, decreases in detection thresholds due to the presence of the enhancing precursor were often 10 dB or more. He considered mechanisms that could possibly be responsible for these large improvements in threshold, suggesting that some form of adaptation of the frequency regions surrounding the target frequency was likely involved. A subsequent study by Viemeister and Bacon (1982; see also recent study by Byrne *et al.*, 2011) helped to clarify this point. They demonstrated that enhanced tones could produce more forward masking than unenhanced tones presented at equal levels. One important implication of this work was that the enhancement effect appeared to act like a gain in the physical stimulus, a finding that would be more consistent, they argued, with adaptation of inhibition than of excitation. This result raised the question of whether central sites of inhibition might contribute to the effect because of the lack of any evidence for such a post-stimulus gain mechanism in the periphery. Physiological work consistent with this speculation, to some extent, was reported by Palmer *et al.* (1995), who found evidence at the level of the auditory nerve in guinea pig for relative changes in response patterns for neural elements

tuned to target and masker frequencies using stimulus conditions similar to those producing enhancement in psychophysical experiments. While their results indicated different amounts of adaptation of the frequency regions representing the target and masker due to the precursor, they did not observe evidence for gain at the target frequency, leaving open the possibility that higher levels of the auditory pathways contributed to the forward masking results of Viemeister and Bacon (1982). Recently, physiological evidence supporting the view that higher-level auditory mechanisms contribute to enhancement has been reported by Nelson and Young (2010) in some units of the inferior colliculus in marmoset monkey. They proposed that a mechanism based on adaptation of inhibition—consistent with Viemeister and Bacon’s (1982) hypothesis about an increase in gain at the target frequency—that develops as the stimulus propagates beyond the auditory periphery could form a neural basis for enhancement. Note here the distinction between differential peripheral adaptation across the frequency regions representing masker and target and “absolute gain” at the target frequency. The former appears to be well established in the auditory nerve, whereas the latter seems to manifest as the stimulus propagates up the auditory pathways.

One potential complication to the conclusion that higher-level auditory mechanisms (above the auditory nerve) contribute to enhancement, however, was the lack of any evidence for enhancement occurring psychophysically when the precursor was presented contralateral to the target-plus-masker (e.g., Viemeister, 1980; Carlyon, 1989; Kidd and Wright, 1994; Serman *et al.*, 2008). If enhancement was due at least in part to mechanisms acting beyond the auditory nerve, and thus at or beyond the point where the inputs from the two ears are combined, then it might be expected that contralateral input of the precursor would provide some degree of benefit. At the very least, a lack of contralateral enhancement would be consistent with, although not conclusive evidence for, a purely peripheral (e.g., cochlear) origin. Conversely, evidence of contralateral enhancement would likely rule out a cochlear site of origin. While it is not necessarily the case that contralateral precursor presentation should activate the same higher-level inhibitory mechanisms as ipsilateral presentation, the lack of any contralateral benefit was perhaps surprising and at the very least provided motivation for closer scrutiny of the issue. The recent physiological study reported by Nelson and Young (2010) did not separate ear of presentation, so, while their findings supported the notion of enhancement comprising some component of gain originating at higher levels in the auditory system, their work did not settle the issue of a possible contralateral contribution to the effect.

The idea that mechanisms beyond the cochlea and auditory nerve could be involved in enhancement received further support in a somewhat different psychophysical paradigm from the work of Richards *et al.* (2004). In their study, a pure-tone target was masked by a highly uncertain multi-tone masker having frequency components drawn randomly from presentation to presentation. Informational masking is thought to originate beyond the auditory periphery, so stimulus manipulations that release informational masking may

originate at higher levels as well. Again, though, that argument is not conclusive because enhancement occurring peripherally could simply attenuate the masker prior to the competition at higher levels thought to be basis for informational masking. Richards *et al.* (2004) found that presenting the masker as a cue prior to the target-plus-masker provided a significant improvement in detectability. Much less benefit was found for other types of cues such as the target-plus-masker as a precursor or the masker as a postcursor. Importantly, Richards *et al.* (2004) found that this “masker-first advantage” could be produced by contralateral presentation, although the magnitude of the benefit was less than when the precursor was ipsilateral and, as is typical in studies where informational masking dominates, there were large individual differences among subjects.

One possibility for the discrepancy between the Richards *et al.* (2004) study and the earlier work on enhancement cited in the preceding text with respect to whether a contralateral precursor was beneficial was that the Richards *et al.* (2004) experiment was conducted under conditions of high masker uncertainty, whereas the earlier work not. The masker in the Richards *et al.* (2004) study was a multitone complex having frequencies chosen at random on every presentation that is known to produce large amounts of informational masking (e.g., Neff and Green, 1987; also review in Kidd *et al.*, 2008b). Thus it is possible that the benefit of the contralateral cue in the Richards *et al.* (2004) experiment was due to a reduction in informational masking resulting from decreased masker uncertainty. For a variety of reasons, including the lesser effectiveness of a target-plus-masker cue preceding the test stimulus, Richards *et al.* (2004) concluded that a reduction in uncertainty alone was insufficient to account for their results (see article for details). However, the exact nature and origin of the contralateral benefit they found is still not clear, and questions generated by their findings form one rationale for the current study.

The contextual effect produced through the masker-first cue reported by Richards *et al.* (2004) does not stand alone in providing evidence supporting a role of binaural mechanisms in enhancement-related studies. Serman *et al.* (2008) conducted an experiment in which an enhancing precursor (notched-filtered noise) created the sensation of pitch in a subsequent broadband pink-noise test stimulus. The idea was that the pitch evoked in the test stimulus was related to the frequency of the notch in the precursor. To measure this effect, a third “probe” stimulus—a narrow band of noise—followed the test stimulus. The task of the listener was to judge whether the probe band was higher or lower in frequency than the notch in the precursor (and thus the pitch evoked in the test stimulus). When the precursor and test stimuli were presented to both ears with the same interaural time difference (ITD; yielding the same apparent interaural location), pitch discrimination was about 84% correct on average. When the precursor and test stimuli were presented to both ears with different ITDs (i.e., different apparent locations), discrimination performance dropped to about 70% correct. In both cases, the precursor and test stimuli were the same in each ear and, if the effect was determined solely by the monaural input, would be expected to produce results equal to the (true monaural)

control. The fact that performance was worse when the precursor and test stimuli differed in ITD led to the conclusion that “enhancement is significantly dependent on binaural processing” (p. 4415). It is not entirely clear why an abrupt change in perceived location of these sequential stimuli should have adversely affected performance. An earlier study by Kidd and Wright (1994) found little decrease in performance under conditions that, superficially at least, appeared similar in that the precursor shifted abruptly in perceived location relative to the masker during the stimulus sequence. In the Kidd and Wright (1994) study, the shift in location of the precursor was produced by an interaural level difference (ILD) rather than an ITD. The substantial differences in stimuli and tasks—discriminating a pitch difference in an enhanced noise band (Serman *et al.* 2008) and detection of a very brief “pure” tone embedded in a notched noise (Kidd and Wright, 1994)—are the likely cause of this apparent discrepancy, although this issue requires further examination.

In addition to the psychoacoustical studies in the preceding text that employed detection or pitch discrimination tasks using relatively simple stimuli, there has been work that seems related to enhancement reported in the speech perception literature, including some that bears directly on the issue of the potential benefit of providing contralateral context. As noted in the preceding text, some of the earliest studies of enhancement employed harmonic complexes as precursor and test stimuli. Summerfield *et al.* (1984, 1987) used a paradigm in which the complement of a vowel spectrum (formants were converted from peaks in the spectrum to troughs) was played immediately before a flat spectrum harmonic complex. They found that the resulting “enhanced” harmonic complex could readily be identified as the correct (the complement of the precursor) vowel. Moreover, the enhanced peaks appeared to be louder than the peaks of the actual vowel (under appropriate controls) lending support to the hypothesis of Viemeister and Bacon (1982) that enhancement produces gain in the auditory pathways. However, contralateral presentation of the precursor did not provide any evidence for improvement in identification performance. Note that this task involved discrimination of suprathreshold stimuli. Thus the work of Summerfield *et al.* (1984, 1987; see also Summerfield and Assman, 1989) was in general agreement with the earlier studies using a detection task that reported no contralateral enhancement effect.

In contrast, there is another series of studies of contextual effects in speech perception that has found evidence for a significant benefit from contralateral stimulation (e.g., Lotto *et al.*, 2003; Holt, 2005). The task that was typically employed required the observer to label a given speech token as belonging to one of two phoneme categories. In this categorical perception experiment, a series of tokens is generated along the stimulus dimension thought to underlie phoneme identity—most often the frequency of a spectral contrast (change in the location of a peak in intensity over time). Typically, there is a sharply defined performance boundary (point where labeling changes from one phoneme to the other) along this stimulus dimension. Holt and colleagues (e.g., Holt *et al.*, 2000; Holt and Lotto, 2002; Holt, 2005, 2006a,b) have shown that the location of the boundary may

be influenced by the presence and spectral composition of a precursor sound. For example, if the phoneme alternatives are /ga/ and /da/, which may be distinguished principally by the onset frequencies of the second and third formants, the concentration of energy in frequency of the precursor exerts a contrastive influence on the categorical responses. If the precursor is relatively high, the responses increase for the lower-frequency-onset stimulus (/ga/), and vice versa. Such a contrast enhancement effect is likely to be utilized by the listener in the perception and processing of connected speech. However, this contrastive emphasis appears to be auditory in origin—rather than linguistic—because nonspeech precursors affect the boundary just as do actual speech precursors. The most important finding for the purposes of the current study is that this shift in category boundary due to the precursor may be produced contralaterally although, as with the Richards *et al.* (2004) finding in the preceding text, it is less effective than for ipsilateral presentation. With respect to the discussion about the task- and stimulus-dependent factors governing enhancement, it should be noted that this effect occurred in the absence of any explicit masker and under conditions of relatively low stimulus uncertainty. Thus contralateral context effects for speech identification—just as with enhancement for tone or noise band detection/discrimination—appears to depend critically on the specifics of the experimental design and stimuli.

The evidence reviewed in the preceding text suggests that contextual effects may serve to emphasize the perception of spectral contrasts in sequences of sounds leading to changes in detectability or masking effectiveness or altering the perceptual boundary between phoneme categories. However, the evidence often appears to be incomplete and even at times points to contradictory conclusions. One hypothesis that emerges from the review of the work in the preceding text on enhancement and possibly related phenomena is that there may be more than one mechanism underlying the observed benefits to performance (cf. Summerfield and Assman, 1989; Richards *et al.*, 2004; also recent study by Erviti *et al.*, 2011). The idea here is that some benefit may be obtained simply through the differential prior stimulation of the frequency regions in which target and masker are presented. Thus the relationship between the precursor and masker only matters to the extent that the precursor adapts the masker frequency regions more than it adapts the target frequency regions. However, if there is another mechanism that contributes to the obtained benefit, then simply providing differential prior stimulation of target and masker frequency regions will not capture the full effect. Should that prove to be the case, it is conceivable that other aspects of the relationship between the precursor and masker influence performance.

The current study was intended to evaluate this hypothesis in a fairly limited way. We sought to design an experiment that would allow us to determine whether differential prior stimulation of the target and masker frequency regions captured the entire enhancement effect. This was accomplished by using two types of precursors that were both intended to differentially adapt (either excitation or inhibition, or both) target and masker frequency regions by

roughly the same amount. However, the two types of precursors varied in the degree to which they corresponded to the spectrotemporal characteristics of the masker. The idea was that one possible mechanism—beyond differential adaptation of target and masker frequency regions—might involve some form of higher-level computation (e.g., taking a difference) between successive sounds. If that were the case, it is possible that this putative additional mechanism would be as effective when the stimuli—precursor and target-plus-masker—were presented to different ears as when they were presented to the same ear, so both ipsilateral and contralateral conditions were examined. The experimental approach we chose used a nonspeech pattern identification task (e.g., Kidd *et al.*, 1998b; Kidd *et al.*, 2002) in which the listener is trained to identify a set of narrowband frequency patterns formed by sequences of tone bursts. This task is appealing because it requires suprathreshold identification of sounds, so, in that sense, has some face validity to situations where contextual effects are likely to be important in realistic listening environments but avoids the complications of the possible role of linguistic processing associated with speech stimuli. It also allows for a high degree of control over the type of masking—energetic or informational—that is present because targets and maskers may be confined to mutually exclusive (or completely overlapping), narrow frequency bands that may be randomized or fixed across trials. Furthermore, the previous work of Richards and Neff (2004) indicated that either a target cue or a masker cue (when presented as precursors) could provide large benefits under highly uncertain conditions. In contrast, the typical low-uncertainty target-cue control condition in enhancement studies (precursor at the target frequency rather than the masker frequency; e.g., Viemeister, 1980; Carlyon, 1989) has yielded little or no advantage. To examine further the possible role of knowledge about the target, a condition was included in the study in which the otherwise random target frequency band was held constant across trials. This manipulation provided a form of target cue² that allowed comparison of the effectiveness of *a priori* information about the masker with *a priori* information about the target.

II. METHODS

A. Listeners

A total of 10 listeners participated in this study.³ The ages ranged from 21 to 27 yr. All of the listeners had normal hearing as determined by standard audiometry.

B. Stimuli

All of the target stimuli were sequences of temporally contiguous pure tones that were arranged into six spectrotemporal patterns falling within a specified narrow frequency band. These patterns are essentially the same as those that have been used in previous studies of nonspeech pattern identification (cf. Kidd *et al.*, 1995; Kidd *et al.*, 1998a; Kidd *et al.*, 1998b; Kidd *et al.*, 2002; see also Weber, 1988) although in the current study, the patterns consisted of only four pure-tone elements rather than eight elements as were

used in the past. The maskers were eight simultaneous sequences of four pure tones randomized in frequency within narrow bands on each trial so that it was highly unlikely that they would form reliable patterns (cf. Kidd *et al.*, 2002). The masker and target elements were presented synchronously throughout the sequence. In all of the conditions described in the following text, the narrow frequency bands occupied by masker tones were also selected at random on each trial.

The narrowband tone sequences—for both targets and maskers—were constructed from 16 frequency bands spaced equally on a logarithmic frequency scale. The center frequencies of the bands were separated by a ratio of about 1.262:1 and ranged from 218 to 7129 Hz. Within each band, there were seven frequencies used to construct the stimuli with each frequency separated by intervals of about 3% spaced symmetrically around the center frequency (e.g., the lowest-frequency band ranged from 200 to 236 Hz). The six target patterns are illustrated schematically in Fig. 1. Among the 16 frequency bands, the target patterns occurred in either band 4, 7, 10, or 13 corresponding to band center frequencies of about 438, 880, 1767, and 3549 Hz. When a target band was selected, the two frequency bands immediately adjacent to the target band on that trial were excluded from containing maskers. This restriction was implemented to reduce energetic masking (e.g., Kidd *et al.*, 2008b) of the target. Thus the maskers were drawn from 8 of the 13 remaining frequency bands on a given trial after selection of the target and exclusionary bands.

Each sequence of tones comprised four pure-tone pulses that were each 60 ms in duration, including 10-ms cosine-squared gating ramps, having starting phases that were chosen randomly on each presentation. For the target sequence, the frequency band containing the target elements was either fixed across trials within a block or was chosen randomly on each trial from the four alternatives listed in the preceding text. For the masker sequences, 8 of the remaining 13 bands

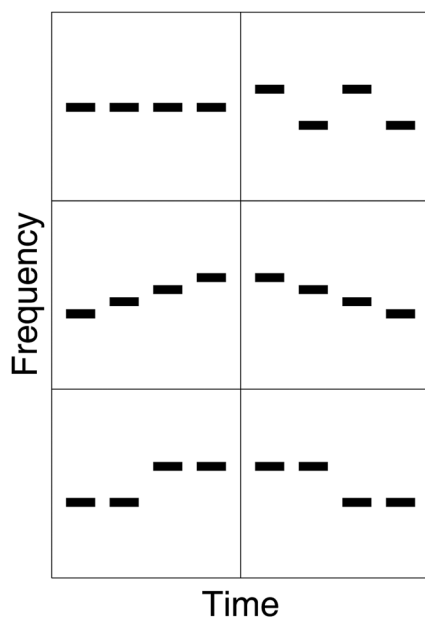


FIG. 1. A schematic illustration of the six target patterns in sound spectrogram format.

were chosen at random on each trial regardless of whether the target band was fixed or chosen at random. Figure 2 illustrates an example of a target (bold lines) embedded in a masker sample along with a preceding “exact masker” cue (stimulus on left; see following text) in sound spectrogram form. The level of each tone in the target was held constant at 60 dB SPL. On every trial the levels of the eight individual masker sequences (see Fig. 2) were randomized within a 20 dB uniform range with each of the four tones within each sequence set to the same level. After level randomization the masker tones were summed and the resulting stimulus was presented at 60 dB SPL. Thus all of the results presented were obtained at a target-to-masker ratio (T/M) of 0 dB.

C. Procedures

The stimuli were generated in MATLAB and presented at a 50-kHz sampling rate through Sennheiser HD 280pro headphones driven by a Tucker-Davis Technologies digital-to-analog converter (System II, 16-bit). Stimulus presentation, response recording and feedback were also implemented via MATLAB. A graphical user interface was used to provide instructions/feedback, display icons of the target patterns, and record responses. The listeners were situated in a sound-attenuating double-walled IAC booth. Except for one condition in which the cues were presented contralaterally, listening was monaural through the right earphone.

The experimental task was one-interval six-alternative forced-choice with the six target patterns forming the response alternatives. There were three 2-h sessions conducted for each subject. The first session (training) familiarized the listeners with the target patterns at all of the possible frequencies and helped them learn to identify the patterns via response feedback. The training session combined fixed- and random-frequency target band conditions in equal proportions. All of the listeners achieved quiet identification scores higher than 90% correct after training. The training phase for quiet target identification lasted for the first hour and was followed by target identification in the presence of the maskers used in the

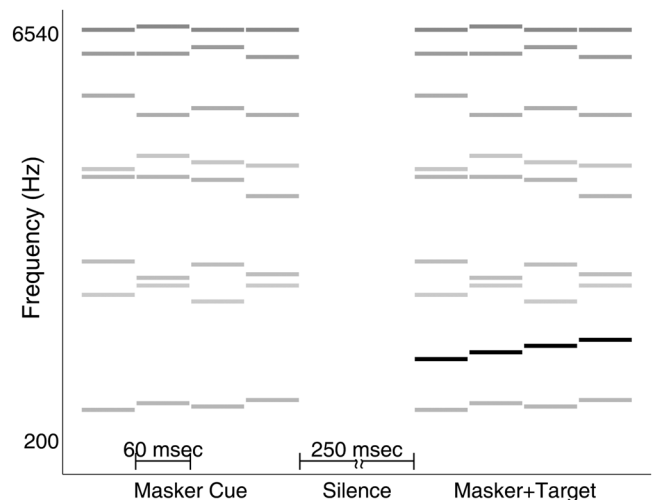


FIG. 2. A schematic illustration in sound spectrogram form of a sample trial containing an exact-masker cue (grays) followed by a target (bold) plus masker (grays). The gray scale indicates higher levels with darker shading.

first test session (described in the following text) again divided equally between fixed- and random-frequency target-band presentations. In both fixed and random conditions, the four target frequency bands were equally represented.

After the training session, two test sessions were completed. The order of the test sessions was the same for every subject. In each session, the control condition (random target-frequency band) and two comparison conditions were tested. In the first test session following training, the two comparison conditions were designated “fixed target-frequency band” and “ipsilateral exact-masker cue.” In the fixed target-frequency band condition, one of the four target frequency bands was selected and held constant throughout the block of trials. In all other conditions in the study, the target frequency band was selected at random on each trial from the four alternatives. In the ipsilateral exact-masker cue condition, a copy of the masker was presented before the target-plus-masker (with 250 ms of silence from the end of the cue to the beginning of the target-plus-masker).

In the second test session, the uncued control condition and two different comparison conditions—called the “notched-noise cue” and “contralateral-masker cue” conditions—were tested. In the notched-noise cue condition, a notched-filtered Gaussian noise was presented prior to the target-plus-masker. The notch was centered on the target band for that trial and was equal in width to the three-band “protected region” around and including the target band. The noise bandwidth equaled the total frequency range of the multitone maskers extending from 200 to 7717 Hz, and the overall level was equal to the masker at 60 dB SPL. In the contralateral-masker cue condition, the cue was an exact copy of the masker—as in the preceding session described in the preceding text—but was presented contralateral to the target-plus-masker. Otherwise, the methods employed in the second test session were identical to those used in the first test session.

A comparison of the results from the two repetitions of the uncued control condition indicated that the listeners’ performance was not significantly different in the two sessions. Thus in the presentation of the results and statistical analyses that follow, the data for the uncued control condition were pooled across the two sessions. The findings from the first test session (uncued control, fixed target-frequency band, and ipsilateral exact-masker cue) will be referred to as forming Experiment 1 while the findings from the second test session (uncued control, notched-noise cue, contralateral-masker cue) will be designated Experiment 2.

III. RESULTS

A. Effects of reducing target-frequency band uncertainty and masker cuing

Figure 3 displays the results of Experiment 1 in which the uncued control condition was contrasted with the fixed target-frequency band and ipsilateral exact-masker cue conditions. The abscissa is the center frequency of the band containing the target and the ordinate is percent correct identification converted to rationalized arcsine units (RAUs, [Studebaker, 1985](#)) because of the wide range of performance observed including values approaching 100% correct. The

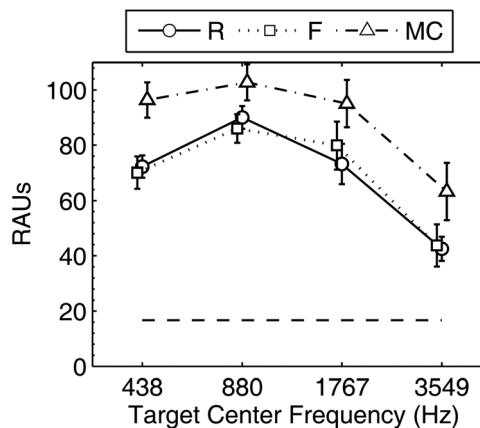


FIG. 3. The group-mean results from the first experiment. The abscissa is the center frequency of the four target bands while the ordinate is percent correct identification performance converted to rationalized arcsine units (RAUs). The error bars are plus and minus one standard error of the means. The results from the uncued random target-frequency band control condition (R) are plotted as circles connected by solid lines; the fixed target-frequency band condition (F) as squares connected by dotted lines; and the ipsilateral exact-masker cue condition (MC) by triangles connected by dashed-dotted lines. The points are slightly offset along the abscissa for clarity. The dashed line near the bottom indicates chance performance (1/6).

values plotted are group means and standard errors for the uncued control condition (circles connected by solid lines, denoted “R”) and both fixed target-frequency band (squares connected by dotted lines, denoted “F”) and ipsilateral exact-masker cue (triangles connected by dashed-dotted lines, denoted “MC”) conditions.

With respect to the question motivating this experiment the answer is clear: Relative to the uncued control condition, fixing the target frequency band did not assist in identification performance. In contrast, providing a preview of the masker helped considerably. Averaged over listeners and the four target band frequencies, performance was 69.5 and 69.9 RAUs for the uncued control and fixed target-frequency band conditions, respectively. For the ipsilateral exact-masker cue, the corresponding value was 89.4 RAUs, about 20 RAUs higher than the uncued control. For each condition tested, the best performance occurred when the target was presented in the lower mid-frequency band centered at 880-Hz (band 7 of 16) and the poorest performance was found for the highest-frequency band centered at 3549 Hz (band 13). The generally poorer pattern identification performance for the higher-frequency target bands has been observed in past studies using similar stimuli (cf. [Kidd *et al.*, 1998b](#); [Kidd *et al.*, 2002](#)). For the uncued control condition, identification performance was about 90.1 and 41.8 RAUs for those two target frequency bands, respectively. The corresponding values were 86.0 and 43.7 RAUs for the fixed target-frequency band condition, and 102.8 and 63.2 RAUs for the ipsilateral exact-masker cue condition. Thus the same qualitative effect of absolute target-frequency band was found in both uncued and cued conditions.

B. Notched-noise and contralateral-masker cues

In the first experiment, an exact copy of the masker preceding the target-plus-masker in each trial provided a

significant cued advantage. In this second experiment, the benefits of two different types of masker cues, notched-noise and an exact-masker cue presented contralaterally, were examined. The group mean results from the second experiment are shown in Fig. 4 plotted in the same manner as Fig. 3. As in the first experiment, the data in percent correct were transformed to RAUs and performance in the uncued control condition is plotted for comparison. Both the notched-noise cue (NC) and the contralateral-masker cue (CC) yielded better scores overall than the uncued control condition (R). The group-mean notched-noise cue condition averaged about 74.9 RAUs while the corresponding value for the contralateral-masker cue was about 79 RAUs. As stated in the preceding text, group mean performance for the uncued control condition was 69.5 RAUs. The trend for poorer overall performance for the highest target frequency band, as well as best performance for the mid-low frequency band, held up for these two cued conditions as well. Interestingly, and in contrast to the ipsilateral exact-masker cue used in Experiment 1, the performance advantage for the highest frequency band was negligible for these two cues.

C. Contextual benefits and individual differences

A two-way repeated-measures analysis of variance was performed on the data from these two experiments using the individual-listener RAU-transformed values. The two experiments were combined such that the main effect of condition included the averaged uncued control, fixed target-frequency band, ipsilateral exact-masker cue, notched-noise cue and contralateral-masker cue conditions. The results indicated that main effects of condition [$F(4,36) = 8.81, P < 0.001$] and target frequency band [$F(3,27) = 22.87, P < 0.001$] were significant as was the interaction [$F(8.5,76.7) = 2.05, P = 0.048$]; using the Huynh-Feldt correction]. To make multiple comparisons of the various cue conditions with the common no-cue control condition, Dunnett's test was used. The results indicated that the fixed target-frequency band condition was not significantly different than the uncued control condition

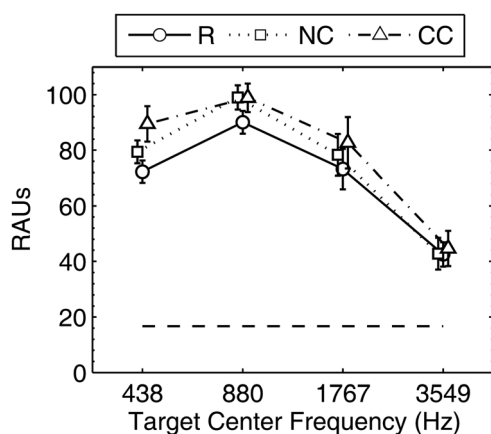


FIG. 4. The results from the second experiment plotted in the same manner as Fig. 3. The results from the uncued random target-frequency band condition are plotted by circles connected by solid lines (R; replotted from Fig. 3); the notched-noise cue condition (NC) by squares connected by dotted lines; and the contralateral-masker cue condition (CC) by triangles connected by dashed-dotted lines.

($P > 0.05$). The comparisons of the ipsilateral exact-masker cue and notched-noise masker cue to the control were significant at the $P < 0.01$ level, and comparison of the contralateral-masker cue condition to the control was significant at the $P < 0.05$ level.

Figure 5 shows the group mean benefits obtained from the four test conditions relative to the uncued control condition averaged across the four target frequencies. Roughly, the identification performance advantages were about 0.5 RAU for the fixed target-frequency band, 19.9 RAUs for the ipsilateral exact-masker cue, 5.3 RAUs for the notched-noise cue, and 9.4 RAUs for the contralateral-masker cue. It is possible that the current results underestimate the benefits provided by the cues in some cases because of ceiling effects where cued performance approached 100% correct. For example, for the best condition, the ipsilateral exact-masker cue, the advantage for the best target frequency band of 880 Hz was only 12.7 RAUs, whereas it was 24.1, 21.9 and 20.7 RAUs at the 438-, 1767-, and 3549-Hz target bands, respectively, where uncued performance was poorer.

Typically, in masking experiments dominated by informational masking, there are large intersubject differences in performance. Figure 6 shows individual results plotted as histograms, indicating the magnitude of the benefit obtained in each test condition relative to the uncued control condition. For each of the 10 subjects, the benefit is shown averaged across the four target frequencies. Note that relative to the uncued control a majority of the subjects (6 of the 10) actually had somewhat poorer performance for the fixed target-frequency band condition while two listeners had substantial benefits (falling in the 15 and 25 RAU difference ranges). For the two exact-masker cues, a majority of the subjects benefitted from the cue with 7 of 10 subjects falling in the 20-RAU advantage range or higher when the cue was ipsilateral and 5 of 10 subjects falling in the 10-RAU benefit or higher categories when the cue was contralateral. In the case of the notched-noise cue, only about one-half of the subjects showed some benefit and none exceeded the 15 RAU range.

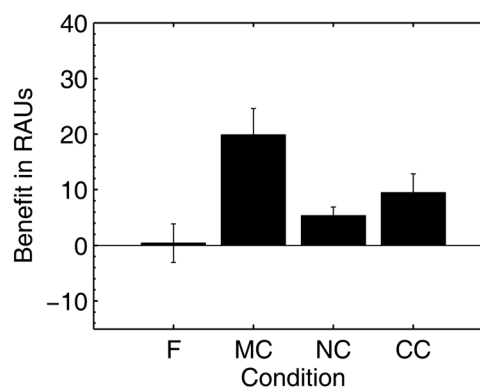


FIG. 5. The advantage due to *a priori* information for the various conditions of both Experiments 1 and 2. The abscissa is the cued condition reading from left to right: fixed target-frequency band (F), ipsilateral exact-masker cue (MC), notched-noise cue (NC), and contralateral-masker cue (CC). The ordinate is group-mean improvement in identification performance expressed in RAUs relative to performance in the uncued random-frequency control condition. The error bars are the standard errors of the group mean differences.

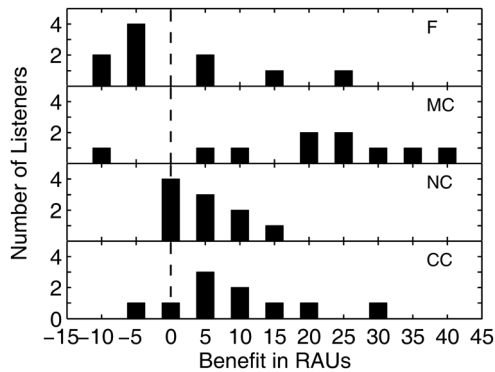


FIG. 6. Histograms of the benefit (re. uncued control) afforded to identification performance (in RAUs) for individual subjects in each of the four test conditions. The abscissa is cued benefit in 5-RAU bins while the ordinate is the number of subjects falling in each range. The conditions are (top to bottom): fixed-target frequency, ipsilateral exact-masker cue, notched-noise cue, and contralateral-masker cue.

IV. DISCUSSION

The current study has found evidence for beneficial contextual effects for the task of nonspeech pattern identification tested under conditions of high informational masking. This work thus extends the body of evidence documenting sequential effects that improve performance by emphasizing spectral contrasts to a new class of stimuli and task. While improved performance was found for both types of precursors tested here, greater improvements occurred when the precursor also matched the spectrotemporal pattern of the subsequent masker. Furthermore, the benefit of a precursor that exactly matched the masker could be obtained by contralateral presentation. No difference in performance was observed between conditions in which the frequency band containing the target was held constant across trials versus randomized across trials, suggesting that, for this nonspeech pattern identification task, target-band frequency uncertainty has little effect on performance. These results are reviewed and discussed in the following text, but overall the conclusion is that these findings cannot be accounted for by a single adaptation-based mechanism and that one or more additional mechanisms or processes contribute to the observed performance benefits.

A. Lack of benefit for certain target-frequency band

The absence of an advantage for holding target-frequency band constant across trials compared to randomizing target-frequency band was somewhat surprising based on past work. First of all, cuing a pure-tone target has been shown to produce modest reductions in masked detection thresholds (about 6-8 dB) when the masker was a multitone complex that varied randomly from presentation to presentation (Neff and Callaghan, 1988; Richards and Neff, 2004). Adding to the overall uncertainty by also randomizing the frequency of the target in highly uncertain masker conditions tends to adversely affect performance even more (cf. Spiegel and Green, 1982; Richards and Neff, 2004; Kidd *et al.*, 2008a) presumably providing a greater opportunity for contextual information to be effective. That is, performance

under stimulus presentation conditions producing large amounts of informational masking may often benefit substantially from manipulations that segregate the target from the masker or that direct attention to the target (cf. Kidd *et al.*, 2008b). Generally, higher amounts of informational masking produce greater potential advantage. These previous studies of reductions in masked detection thresholds when target cuing was provided in highly uncertain listening conditions are at odds with the current findings for pattern identification in similarly uncertain listening conditions.

At present it is not clear why varying target-frequency band uncertainty should produce no effect for suprathreshold pattern identification but significant—and sometimes quite large—effects for different tasks as reported in other studies. One possibility is that the target-band randomization employed here was not sufficient to produce enough of an increase in uncertainty beyond that caused by the masker randomization to allow cuing the target to have a demonstrable effect. Taking into account the variability both within and across frequency bands, the number of possible maskers on any given stimulus presentation is extremely large. In the current experiment, fixing the target frequency band reduced the number of alternative target frequency band regions from four to one and the number of possible targets from 24 to 6 (recalling that there were six possible target patterns for any frequency region). Thus considering the detection studies cited in the preceding text, that also had very large numbers of potential maskers and thus very high masker uncertainty, the manipulation here of fixing the target frequency band region fell in between the Richards and Neff (2004) study and the Kidd *et al.* (2008a) study with respect to the reduction in uncertainty as gauged by the reduction in the number of alternative targets. So it does not seem likely that the different findings are related solely or simply to differences in the amount of reduction in target frequency uncertainty. None of the fixed-target cases in the preceding text reduced the overall stimulus uncertainty in any meaningful way when considered in terms of the total number of possible target-plus-masker combinations that was dominated by the very large number of possible maskers.

Other possibilities for the lack of an identification performance advantage from reducing target-frequency band uncertainty here include the above-threshold nature of the discrimination task, as opposed to the detection studies discussed in the preceding text. We have not measured psychometric functions under fixed- and random-target frequency band conditions, so we do not currently know whether a benefit might be apparent as the stimuli are lowered in level to T/M values near detection threshold. Also, there is target-frequency uncertainty, in one sense, even in fixed target-frequency band presentation because of the one-in-six random selection among the target set inherent to this pattern identification task. Solving the task in the uncertain-frequency band condition requires that 24 absolute frequency patterns be remembered accurately or, more likely, adopting an analysis strategy based on relative frequency in which the variation in absolute frequency is normalized. Thus six patterns of relative frequency could be stored and recalled in either fixed or random target-frequency band conditions

effectively equating uncertainty. As has been remarked upon in previous articles describing experiments in which this nonspeech pattern identification paradigm was employed (cf. Kidd *et al.*, 1998b; Kidd *et al.*, 2002), the target pattern seems to “emerge” perceptually over time as the burst sequence progresses.⁴ Holding in memory a specific frequency marker for the beginning of a target pattern or explicitly directing attention to a specific absolute frequency region does not seem to assist in the processing of the entire complex, time-varying stimulus. The sequential nature of the pattern recognition task in which the target frequencies are compared with stored templates of the frequency relations forming the patterns, as proposed in the preceding text, does not appear to benefit from knowledge of the starting point of the pattern. Identification only occurs when sufficient evidence has accumulated during the burst sequence for a pattern match. The current pattern identification procedure does not allow for explicit cuing with an exact target precursor, but informal listening to a variety of nontarget precursors⁵ has yielded no indication of any target-cue advantage. However, the limited study of this problem and the rather large discrepancy between the current findings and those of the relevant detection experiments suggest that further study of the role of target frequency uncertainty in the different tasks is needed to resolve this issue.

One further point about absolute frequency of the target band should be considered in light of the current findings. A pervasive effect on performance was found as a function of target band center frequency. The most obvious manifestation of this effect was at the highest target band frequency where identification performance was poorest in every condition tested. This effect does not appear to be related to the presence of target band frequency uncertainty per se because it occurred in every condition including the certain target-frequency band case. The most likely explanation for this frequency effect is a mismatch between the rule used to scale the relative frequencies of the target elements as absolute frequency of the target band varied (constant proportion of band center frequency) and the perceptual ability to distinguish the variations signaling the different target patterns. Some earlier evidence from our laboratory is consistent with this view. Kidd *et al.* (1998a) found that pattern identification improved at fixed T/Ms as the frequency range from which the patterns were constructed was increased over the range tested which extended to about 50% of the center frequency. The earlier findings suggest that the frequency effect found here could be reduced or eliminated by exaggerating the frequency differences among target pattern elements as center frequency increases.

B. Notched-noise advantage

The small but significant cued advantage found for the notched-filtered noise poses some difficulties for explanations of the advantage based solely on a reduction in masker uncertainty. Although the power spectra of the noise cue and multitone masker have similar frequency ranges, masker uncertainty is unaffected by cuing with the notched noise. However, the presence of the gap in the noise spectrum

centered on the target frequency region means that the neural elements representing the target and masker are differentially affected prior to the test stimulus. As discussed in Sec. I, differential prior stimulation of target and masker frequency regions may, under certain conditions, lead to enhancement in which the effective subsequent response of the target region is increased relative to the masker regions (Viemeister, 1980). It seems reasonable to think that the current notched-noise cue advantage is another example of the same enhancement effect, this time acting to improve the identification of suprathreshold, narrow-band, time-varying nonspeech patterns. The small increase in percent correct performance could reflect a significant gain in target level given the very shallow slopes of the performance-level functions reported in past studies using a similar paradigm (e.g., Kidd *et al.*, 1998b; Kidd *et al.*, 2002). Based on studies describing the relation between changes in masker level and masked threshold for analogous pure-tone forward masking results (Widin and Viemeister, 1979; Kidd and Feth, 1981), Viemeister and Bacon (1982) estimated that the enhancement effect increased the effective target level by as much as about 15 dB in some cases. We did not obtain psychometric functions in the current study, and that much of a benefit here seems unlikely (cf. Kidd *et al.*, 2002).

The greater advantage found for the ipsilateral exact-masker cue than the noise cue suggests that it is not simply the degree of prior stimulation of the ipsilateral masker channels, and lack of stimulation of the target channel, that fully explains the cued benefit. If that was so, then only the energy of the cue in the masker channels and not the organization of the time-frequency elements of the cue would determine cue effectiveness. While the notched-noise and ipsilateral exact-masker cues were presented at the same overall level, there were differences in the spectrotemporal structure of the cues (i.e., flat noise versus randomized tone levels within the passbands; see Sec. II) that may have influenced the obtained contextual benefits in ways that currently are not known. The simplest assumption is that the tones in the exact-masker cue would be expected to produce about as much enhancement as an equivalent flat-spectrum noise based on comparisons of other temporal effects for noise and tones related to adaptation, such as forward masking. However, such comparisons are complex (cf. Moore and Glasberg, 1983; Neff, 1986; Savel and Bacon, 2003) and indirect, and so this issue should be examined specifically. Carlyon (1989) concluded that enhancement was relatively insensitive to the level of the precursor, finding about the same benefit for enhancer levels varied over a 30-dB range. Also, according to the limited evidence available, no contralateral benefit has been found for enhancers that were noise (e.g., Viemeister, 1980; Viemeister and Bacon, 1982; Kidd and Wright, 1994; Serman *et al.*, 2008). Here a significant but small (about 5 RAUs) benefit was found for ipsilateral presentation of a notched-noise cue, and, in the cases where non-noise precursors did yield contralateral benefits (e.g., Richards *et al.*, 2004; Holt, 2005), the benefit was substantially less than was found from ipsilateral presentation. Little if any benefit then would be expected from presenting a notched-noise precursor (either ipsilaterally or

contralaterally) to a multitone masker due to grouping via similarity or, as noted in the preceding text, to a reduction in masker uncertainty. A possible exception to the conclusion about the benefit of the notched-noise cue being limited to spectrotemporal enhancement (as opposed to reduction in uncertainty or similarity) is that the frequency of the notch could point the listener toward the frequency region of the target, assisting in the focus of attention on that region. This does not seem very likely though because none of our efforts at finding an effective target cue, including fixing the target-frequency band in this study, have been successful.

C. Masker-cue advantage for contralateral presentation

The explanation for the cued advantage found for contralateral presentation of the masker cue preceding the target-plus-masker is fairly complicated. Part of the complication has to do with the extent to which the consequence of separating the cue and target-plus-masker by ear is revealing with respect to the physiological mechanism forming the basis of the phenomenon. Dichotic presentation of the stimuli serves to isolate the cue, to some degree, from the target-plus-masker representation in the ascending auditory nerve pathway. Evidence for a contralateral cued advantage then would logically rule out a mechanism based on adaptation of the sensory cells of the ipsilateral cochlea and the associated afferent auditory nerve supply. Such evidence often forms the basis for attributing masking effects to peripheral versus central sites of origin.⁶ So, the fact that a significant benefit to performance under the contralateral cue condition was found in the current study argues against a purely ipsilateral adaptation-based explanation for the entire cued advantage.

The limited evidence available from physiological studies specifically searching for a correlate of enhancement (e.g., Nelson and Young, 2010), especially in the auditory periphery (e.g., Palmer *et al.*, 1995; also Holt and Rhode, 2000, in cochlear nucleus for contextual influences in speech perception), is not conclusive with respect to this issue. Summerfield *et al.* (1987) raised the possibility that the difference between the effectiveness of ipsilateral and contralateral precursors in producing enhancement, at least for the harmonically related sounds used in their study, could be due to a difference in the extent to which the stimuli are grouped together. The idea is that ipsilateral presentation of the precursor would promote perceptual grouping with the masker more strongly than contralateral presentation due to the difference in perceived location of the two sounds (e.g., Carlyon, 1989; Kidd and Wright, 1994; Serman *et al.*, 2008). The target would then tend to stand out from the grouped precursor-masker complex more strongly than from the segregated precursor-masker complex. Extending that argument to the current conditions, the ipsilateral exact-masker cue and subsequent masker could group perceptually, with the dissimilar target pattern tending to segregate from the complex. Dichotic presentation would reduce the extent to which the precursor and masker were grouped together thereby reducing the perceived contrast formed by the

target. The present study was not specifically designed to evaluate this hypothesis, and our results are not conclusive regarding perceptual grouping/segregation as a contributing factor. However, the finding of a weaker benefit from contralateral precursor presentation than ipsilateral presentation is consistent with that interpretation.

As with the current findings, Richards *et al.* (2004) reported a more robust benefit when the exact-masker cue preceded the target-plus-masker in the ipsilateral ear than in the contralateral ear. In contrast to the earlier studies of enhancement cited in the preceding text that found no contralateral benefit of context, both the Richards *et al.* (2004) study and the current study employed conditions producing a much greater degree of masker uncertainty. It is possible then that the essential difference in these findings was the extent to which the masker cue acted to reduce masker uncertainty. This would imply that the contralateral advantage is only present when masker uncertainty is relatively high. Although reduction in uncertainty may well be a factor, it should be noted that Richards *et al.* (2004) concluded that that interpretation was not sufficient to account for all of the cued advantage observed in their study. In particular, the large cued advantage for a masker precursor, coupled with the lack of advantage for a target-plus-masker precursor, supported that conclusion. That is, a target-plus-masker cue should reduce uncertainty at least as much as a masker-only cue and, based purely on reduction in uncertainty, should be as effective. Empirically, however, that was not the case. The effectiveness of the masker cue also was sensitive to whether the target frequency was fixed or random, though, and a postmasker cue was significantly less effective than a premasker cue—neither finding being consistent with a simple explanation based solely on reduction in masker uncertainty. Thus, while Richards *et al.* (2004) concluded that the masker-first advantage was not based wholly on peripheral processes, they also concluded that a reduction in the degree of stimulus uncertainty was not sufficient to account for the results.

One possibility for a contralateral-masker cue advantage that is not based explicitly on grouping and segregation, or reduction in listener uncertainty, is the concept of masker minimization proposed by Durlach *et al.* (2003; also see earlier work on the equalization-cancellation model, Durlach, 1972). Durlach *et al.* (2003) contrasted two observer strategies that could lead to enhanced performance but that, in this case, would likely be based on some form of top-down process. In their “Listener-Minimization” strategy, the listener essentially creates an inverse filter based on prior knowledge about the sound to be nulled or canceled (the masker). While this idea as it applies to the current study is simply conceptual at this point, it is consistent with the findings and deserves further examination. Richards *et al.* (2004) also concluded that their findings of a masker-first advantage could be considered, at least in part, consistent with such a conceptual framework.

With respect to the relevant findings from the speech perception literature, summarized in Sec. I, both the current findings and the speech context results demonstrate similar but weaker effects when the enhancer/precursor is presented

contralaterally than when it is presented ipsilaterally. However, in the speech perception experiments, contralateral context effects have been observed when the precursor does not appear to appreciably reduce uncertainty in the experiment. When nonspeech precursors were used there was often considerable stimulus variability present (i.e., sequences of tones chosen randomly from a specified distribution on a trial-by-trial basis), but that does not appear to be essential to producing the observed contextual effects. The extent to which informational masking plays a role in those studies, if at all, is thus not clear. Furthermore, it should be emphasized that the findings of this study are for an identification task, and the speech context effects reported in the preceding text were found in a categorization task, and while the phenomena exhibit some similar stimulus dependencies, the evidence is not strong enough to conclude that they share a common origin.

D. Ipsilateral exact-masker cue: Multiple mechanisms?

The ipsilateral presentation of an exact copy of the masker as a cue provided an advantage of about 20 RAUs in identification performance relative to the uncued control condition. Based on the discussion in the preceding text, there appear to be at least two factors contributing to the benefit provided by this cue. The following discussion is based on the hypothetical role of these two factors, but it is acknowledged that there are currently significant gaps in our knowledge of these phenomena, and inconsistencies in the available evidence, that limit the extent to which general conclusions may be drawn. First, from the notched-noise cue advantage that was found here, some benefit may arise simply from the differential prior stimulation of target and masker channels. This benefit appears to be similar to the general perceptual process that enhances spectral contrasts in sequences of sounds due to lateral suppression (e.g., [Houtgast, 1974](#); [Shannon, 1976](#); however, see also [Thibodeau, 1991](#)). Thus regardless of the spectrotemporal correspondence of cue and masker, simply the fact that the masker channels received prior stimulation and the target channel did not may have enhanced the neural representation of the target channel in the subsequent target-plus-masker stimulus. Our data do not allow us to distinguish between adaptation of excitation of the masker channels and gain in the target channel through adaptation of inhibition. Second, the exact correspondence of cue and masker also appears to have contributed to the cued benefit, either through a reduction in uncertainty about the masker or through some more explicit computational processing, for example, by taking a difference between cue and target-plus-masker. Carrying forward with the speculation about these two hypothetical mechanisms and how they are reflected in the current results, to the extent that the notched-noise cue indicates the benefit of differential prior stimulation of target and masker channels, and the contralateral-masker cue indicates the contribution of an additional mechanism, then this hypothetical second process or mechanism was more effective under the conditions tested in the present study.

This supposition leads to the question of whether there is any evidence from the current results suggesting that the effect of the ipsilateral exact-masker cue advantage is simply the sum of the actions of these two putative processes: enhancement through differential prior stimulation of target and masker frequency regions and the comparison of the spectrotemporal correspondence of cue and masker. Overall, the summed advantage for notched-noise and contralateral-masker cues amounted to about 15 RAUs compared to the 20-RAU advantage for the ipsilateral exact masker cue. However, on an individual subject basis (averaged across the four target frequencies), the correlation between ipsilateral exact-masker cue and the sum of the notched-noise the contralateral cues was not statistically significant ($r=0.43$). Thus it appears that while both putative mechanisms may have contributed to the ipsilateral exact-masker cue advantage found here, it is likely that other factors influenced this advantage as well.

V. CONCLUDING REMARKS

The present study has demonstrated that cued advantages may be obtained for the task of pattern identification of nonspeech targets presented in highly uncertain maskers. The nature and composition of the maskers was intended to produce primarily informational masking. The greatest benefit afforded by cuing was found for an exact copy of the random-frequency multitone masker presented just prior to, and in the same ear as, the target and masker. Smaller but statistically significant cued advantages were also found for a masker cue to the opposite ear and for a notched-filtered noise having a similar frequency range as the masker presented to the ipsilateral ear.

The cued advantages found in this study were compared to seemingly similar phenomena described in the psychoacoustic and speech perception literature as “auditory enhancement” or “context effects.” These phenomena appear to be similar in that they act to emphasize spectral contrasts in sequences of sounds, raising the possibility that one or more general perceptual processes contribute to—but may not fully account for—all three. The processes that we speculate contribute to the cued advantages found in the current study, at least, are adaptation of excitation or inhibition, or both, and comparison of the spectrotemporal correspondence of the masker and cue. Despite the apparent similarities among these phenomena, there are some perplexing differences as well. First of all, the presence of a contralateral cued advantage here is qualitatively like that reported for nonspeech context effects in phoneme categorization (e.g., [Lotto et al., 2003](#); [Holt, 2005](#)) but stands in contrast to the lack of contralateral enhancement found for noise or harmonic complexes for detection/discrimination tasks ([Viemeister, 1980](#); [Viemeister and Bacon, 1982](#); [Summerfield et al., 1987](#); [Summerfield and Assman, 1989](#); [Kidd and Wright, 1994](#)). The contralateral cue benefit in this study could act through a reduction in masker uncertainty (cf. [Richards et al., 2004](#)) but that does not match the conditions under which phoneme category shifts are observed that are not measured in masked conditions at all. Furthermore, if enhancement develops or is

augmented as the stimulus propagates up the ascending auditory pathways, as has been suggested from some of the physiological work on the topic (cf. Palmer *et al.*, 1995; Nelson and Young, 2010), it is unclear why contralateral presentation of notched-noise or harmonic complex enhancers should be wholly ineffective as was found by the earlier perceptual work reviewed in the preceding text. Nor is it clear why differences in the apparent locations of precursor and target/masker, without disrupting the monaural stimulus input, should reduce the benefit of context in some cases (Serman *et al.*, 2008) but not others (Kidd and Wright, 1994). Second, if the benefit found in this study beyond that likely attributable to the “enhancement” component is due to reduction in uncertainty, why are similar effects found for ipsilateral and contralateral context in speech categorization experiments where uncertainty is low and the context does little to affect uncertainty? Finally, the evidence describing the persistence of contextual effects varies widely across studies and specific experimental stimuli and procedures. (e.g., one or more seconds; cf. Viemeister, 1980; Holt, 2005; also related, Zwicker, 1964). These disparate findings have at least two implications: First, it is possible that the different time frames over which the effects persist reflect the actions of independent mechanisms. This is consistent with the speculation here that at least two factors contribute to the observed ipsilateral exact-masker cue advantage. Second, the longer time frames have implications for interactions in sequences of sounds beyond what has typically been studied in the laboratory but that correspond to natural conditions such as connected discourse or conversation. The influence of intervening sounds and the cumulative effect of spectrotemporal contrasts in the perception of streams of sounds, especially in multisource environments, are not well understood. Overall, while each of the effects considered in the preceding text likely plays a role in normal communication, a comprehensive explanation of contextual effects is presently rather elusive.

ACKNOWLEDGMENTS

This work was supported by Grant Nos. DC004545, DC02012, DC010058, DC00100, and DC004663 from NIH/NIDCD and Grant No. FA9550-08-1-0424 from AFOSR. The authors are grateful to Nathaniel I. Durlach, Neal F. Viemeister, and Kyle P. Walsh for comments on an earlier version of this manuscript as well as to the Associate Editor, Laurent Demany, and two anonymous reviewers for their suggestions during the review process.

¹The term “cue” may not be strictly appropriate if it implies that the benefit the stimulus confers involves top-down, voluntary processes rather than bottom-up, automatic processes. The way that this distinction applies to the results obtained under the current experimental conditions is a topic of discussion, so the term “cue” sometimes is used in a more neutral way and simply means a precursor sound intended to aid performance.

²Under certain conditions, fixing the value of a stimulus across trials may be considered to act as an implicit “cue” to performance because of the opportunity it provides to develop a standard that is held in memory for comparison to subsequent test stimuli. An explicit cue presented immediately before the test stimulus also provides a reference that is held in memory—arguably in a different form of memory acting over a briefer time scale—for comparison with subsequent stimuli.

³An 11th subject was also tested. However, her data were not included in the results because of at-ceiling performance in the random (reference) condition. Interestingly, we did test her on the full set of conditions and repetitions at a level of -20 dB T/M. At that target level, her results were quite similar to the group mean data shown in the figures with the magnitude of the three masker-cue advantages among the largest found in the group.

⁴The “emergence” of the target under random-frequency band conditions and high masker uncertainty is an assertion based upon the authors’ subjective impressions. However, logically, if one considers the evidence in each frequency band as a means of testing the six hypotheses (six target patterns), there is little evidence available from the first burst, more after the second burst, etc., generally consistent with the idea that the hypotheses are evaluated as the evidence accumulates during the observation interval. In the current stimulus design there are only four elements in each sequence, but the principle is the same (cf. Kidd *et al.*, 2002).

⁵We informally tested a number of target-only cues in an attempt to produce a cued advantage. These manipulations included presenting a tone or tone sequence at the target frequency prior to the trial and even extending the frequency pattern to the temporal region immediately prior to the stimulus interval; i.e., a longer target pattern beginning before the masker onset. None of these manipulations were effective in producing a significant cued advantage, and the fixed target-frequency band condition tested here was, we believe, representative of these various (ineffective) target-cue manipulations.

⁶Under certain conditions related to those tested here, contralateral stimulation may affect ipsilateral masking as revealed, for example, in studies of physiological or psychophysical estimates of “overshoot” (cf. Kawase *et al.*, 1993; Backus and Guinan, 2006; Turner and Doherty, 1997; Bacon and Liu, 2000; Walsh *et al.*, 2010). Given the very different stimulus parameters and procedures used in these studies from the current experiments, it is not possible at present to determine whether this effect has any direct bearing on the findings of contralateral context effects considered here.

- Backus, B. C., and Guinan, J. J., Jr. (2006). “Time-course of the human medial olivocochlear reflex,” *J. Acoust. Soc. Am.* **119**, 2889–2904.
- Bacon, S. P., and Liu, L. (2000). “Effects of ipsilateral and contralateral precursors on overshoot,” *J. Acoust. Soc. Am.* **108**, 1811–1818.
- Byrne, A. J., Stellmack, M. A., and Viemeister, N. F. (2011). “The enhancement effect: Evidence for adaptation of inhibition using a binaural centering task,” *J. Acoust. Soc. Am.* **129**, 2088.
- Carlyon, R. P. (1989). “Changes in the masked thresholds of brief tones produced by prior bursts of noise,” *Hear. Res.* **41**, 223–236.
- Durlach, N. I. (1972). “Binaural signal detection: equalization and cancellation theory,” in *Foundations of Modern Auditory Theory, Vol. 2*, edited by J. V. Tobias (Academic Press, New York), Chap. 10, pp. 369–462.
- Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Shinn-Cunningham, B. C., and Colburn, H. S. (2003). “Note on informational masking,” *J. Acoust. Soc. Am.* **113**, 2984–2987.
- Erviti, M., Semal, C., and Demany, L. (2011). “Enhancing a tone by shifting its frequency or intensity,” *J. Acoust. Soc. Am.* **129**, 3837–3845.
- Freyman, R. L., Balakrishnan, U., and Helfer, K. (2004). “Effect of number of masking talkers and auditory priming on informational masking in speech recognition,” *J. Acoust. Soc. Am.* **115**, 2246–2256.
- Holt, L. L. (2005). “Temporally nonadjacent nonlinguistic sounds affect speech categorization,” *Psychol. Sci.* **16**, 305–312.
- Holt, L. L. (2006a). “Speech categorization in context: Joint effects of speech and nonspeech precursors,” *J. Acoust. Soc. Am.* **119**, 4016–4026.
- Holt, L. L. (2006b). “The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization,” *J. Acoust. Soc. Am.* **120**, 2801–2817.
- Holt, L. L., and Lotto, A. J. (2002). “Behavioral examinations of the level of auditory processing of speech context effects,” *Hear. Res.* **167**, 156–169.
- Holt, L. L., Lotto, A. J., and Kluender, K. R. (2000). “Neighboring spectral content influences vowel identification,” *J. Acoust. Soc. Am.* **108**, 710–722.
- Holt, L. L., and Rhode, W. (2000). “Examining context-dependent speech perception in the chinchilla cochlear nucleus,” presented at the *23rd Annual Midwinter Research Meeting of the Association for Research in Otolaryngology*, St. Petersburg, FL.
- Houtgast, T. (1974). “Lateral suppression in hearing: A psychophysical study of the ear’s capability to preserve and enhance spectral contrasts,” Ph.D. dissertation, Academische Pers B.V., Amsterdam.

- Kawase, T., Delgutte, B., and Liberman M. C. (1993). "Antimasking effects of the olivocochlear reflex. II. Enhancement of auditory-nerve response to masked tones," *J. Neurophysiol.* **70**, 2533–2549.
- Kidd, G., Jr., and Feth, L. L. (1981). "Patterns of residual masking," *Hear. Res.* **5**, 49–67.
- Kidd, G., Jr., Mason, C. R., and Arbogast, T. L. (2002). "Similarity, uncertainty and masking in the identification of nonspeech auditory patterns," *J. Acoust. Soc. Am.* **111**, 1367–1376.
- Kidd, G., Jr., Mason, C. R., and Chiu, P. C.-Y. (1998a). "Identification of brief auditory patterns," in the *Proceedings of the 16th International Congress on Acoustics and the 135th meeting of the Acoustical Society of America*, 2349–2350.
- Kidd, G., Jr., Mason, C.R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (2008a). "Informational masking," in *Auditory Perception of Sound Sources*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer Science+Business Media, LLC, New York), pp. 143–190.
- Kidd, G., Jr., Mason, C. R., and Rohtla, T. L. (1995). "Binaural advantage for sound pattern identification," *J. Acoust. Soc. Am.* **98**, 1977–1986.
- Kidd, G., Jr., Mason, C. R., Rohtla, T. L., and Deliwala, P. S. (1998b). "Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns," *J. Acoust. Soc. Am.* **104**, 422–431.
- Kidd, G., Jr., Richards, V. M., Mason, C. R., Gallun, F. J., and Huang, R. (2008b). "Informational masking increases the costs of monitoring multiple channels," *J. Acoust. Soc. Am.* **124**, EL223–229.
- Kidd, G., Jr., and Wright, B. A. (1994). "Improving the detectability of a brief tone in noise using forward and backward masker fringes: Monotic and dichotic presentations," *J. Acoust. Soc. Am.* **95**, 962–967.
- Lotto, A. J., Sullivan, S. C., and Holt, L. L. (2003). "Central locus for nonspeech context effects on phonetic identification," *J. Acoust. Soc. Am.* **113**, 53–56.
- Moore, B. C. J., and Glasberg, B. R. (1983). "Growth of forward masking for sinusoidal and noise maskers as a function of signal delay; implications for suppression in noise," *J. Acoust. Soc. Am.* **73**, 1249–1259.
- Munson, W. A., and Gardner, M. B. (1950). "Loudness patterns—A new approach," *J. Acoust. Soc. Am.* **22**, 177–190.
- Neff, D. L. (1986). "Confusion effects with sinusoidal and narrow-band noise forward maskers," *J. Acoust. Soc. Am.* **79**, 1519–1529.
- Neff, D. L., and Callaghan, B. P. (1988). "Effective properties of multicomponent simultaneous maskers under conditions of uncertainty," *J. Acoust. Soc. Am.* **83**, 1833–1838.
- Neff, D. L., and Green, D. M. (1987). "Masking produced by spectral uncertainty with multicomponent maskers," *Percept. Psychophys.* **41**, 409–415.
- Nelson, P. C., and Young, E. D. (2010). "Neural correlates of context-dependent perceptual enhancement in the inferior colliculus," *J. Neurosci.* **30**, 6577–6587.
- Palmer, A. R., Summerfield, Q., and Fantini, D. A. (1995). "Responses of auditory nerve fibers to stimuli producing psychophysical enhancement," *J. Acoust. Soc. Am.* **97**, 1786–1799.
- Plack, C. J., and Oxenham, A. J. (1998). "Basilar-membrane nonlinearity and the growth of forward masking," *J. Acoust. Soc. Am.* **103**, 1598–1608.
- Richards, V. M., Huang, R., and Kidd, G., Jr. (2004). "Masker-first advantage in cues for informational masking," *J. Acoust. Soc. Am.* **116**, 2278–2288.
- Richards, V. M., and Neff, D. L. (2004). "Cuing effects for informational masking," *J. Acoust. Soc. Am.* **115**, 289–300.
- Savel, S., and Bacon, S. P. (2003). "Effectiveness of narrow-band versus tonal off-frequency maskers," *J. Acoust. Soc. Am.* **114**, 380–385.
- Serman, M., Semal, C., and Demany, L. (2008). "Enhancement, adaptation, and the binaural system," *J. Acoust. Soc. Am.* **123**, 4412–4420.
- Shannon, R.V. (1976). "Two-tone unmasking and suppression in a forward masking situation," *J. Acoust. Soc. Am.* **59**, 1460–1470.
- Sheldon, S., Pichora-Fuller, M. K., and Schneider, B. A. (2008). "Priming and sentence context support listening to noise-vocoded speech by younger and older adults," *J. Acoust. Soc. Am.* **123**, 489–499.
- Spiegel, M. F., and Green, D. M. (1982). "Signal and masker uncertainty with noise maskers of varying duration, bandwidth and center frequency," *J. Acoust. Soc. Am.* **71**, 1204–1210.
- Studebaker, G. A. (1985). "A 'rationalized' arcsine transform," *J. Speech Hear. Res.* **28**, 455–462.
- Summerfield, Q., and Assman, P. F. (1989). "Auditory enhancement and the perception of concurrent vowels," *Percept. Psychophys.* **45**, 529–536.
- Summerfield, A. Q., Haggard, M. P., Foster, J., and Gray, S. (1984). "Perceiving vowels from uniform spectra: Phonetic exploration of an auditory after-effect," *Percept. Psychophys.* **35**, 203–213.
- Summerfield, A. Q., Sidwell, A., and Nelson, A. (1987). "Auditory enhancement of changes in spectral amplitude," *J. Acoust. Soc. Am.* **81**, 700–708.
- Thibodeau, L. (1991). "Performance of hearing-impaired persons on auditory enhancement tasks," *J. Acoust. Soc. Am.* **89**, 2843–2850.
- Turner, C. W., and Doherty, K. A. (1997). "Temporal masking and the 'active process' in normal and hearing-impaired listeners," in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Hillsdale, NJ), pp. 387–396.
- Walsh, K. P., Pasanen, E. G., and McFadden, D. (2010). "Overshoot measured physiologically and psychophysically in the same human ears," *Hear. Res.* **268**, 22–37.
- Viemeister, N. F. (1980). "Adaptation of masking," in *Psychophysical, Physiological and Behavioral Studies in Hearing*, edited by G. van den Brink and F. A. Bilten (Noordwijkerhout, The Netherlands), pp. 190–198.
- Viemeister, N. F., and Bacon, S. P. (1982). "Forward masking by enhanced components in harmonic complexes," *J. Acoust. Soc. Am.* **71**, 1502–1507.
- Weber, D. L. (1988). "Detection and recognition of auditory patterns," *Percept. Psychophys.* **46**, 1–8.
- Widin, G. P., and Viemeister, N. F. (1979). "Intensive and temporal effects in pure-tone forward masking," *J. Acoust. Soc. Am.* **66**, 386–395.
- Zwicker, E. (1964). "Negative afterimage in hearing," *J. Acoust. Soc. Am.* **36**, 2413–2415.