The effect of frequency cueing on the perceptual segregation of simultaneous tones: Bottom-up and top-down contributions

Yi Shen^{a)}

Department of Speech and Hearing Sciences, Indiana University Bloomington, Bloomington, Indiana 47405, USA

(Received 1 June 2016; revised 30 September 2016; accepted 8 October 2016; published online 8 November 2016)

Listeners were presented with two simultaneous tones of different frequencies (more than one octave apart) and asked to identify the tone that was amplitude-modulated while a tonal precursor was presented to cue the frequency of the lower frequency tone. Performance thresholds were estimated based on the duration of the tone-pair. In Exp. I the duration of the precursor varied from 100 to 400 ms and the inter-stimulus interval (ISI) between the precursor and the tone-pair varied from 0 to 1 s. The presence of the precursor facilitated segregation. As the ISI increased, the facilitation effect of the precursor increased for the precursor durations of 100 and 200 ms, but not for the 400-ms precursor duration. When the precursor was presented to the contralateral ear relative to the tone-pair in Exp. II, no significant change to the precursor effect was observed. These observations contradict the predictions of the model based solely on bottom-up processing, suggesting the likely involvement of top-down processes. © 2016 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.4965969>]

[EAS]

Pages: 3496–3503

I. INTRODUCTION

Two simultaneously presented acoustic stimuli can be perceived either as segregated auditory objects or a single fused auditory object. Many previous studies on the segregation of simultaneous sound sources have identified a number of acoustic cues that influence the perceptual organization, including temporal onset/offset asynchrony (e.g., Bregman and Pinker, 1978; Dannenbring and Bregman, 1978; Rasch, 1978), harmonicity (recently reviewed by Micheyl and Oxenham, 2010a), and sequential contexts (e.g., Darwin *et al.*, 1989; Darwin and Hukin, 1997). On the other hand, perceptual segregation is not completely driven by the stimulus and may be subject to top-down modulations (e.g., van Noorden, 1975). For example, identical stimuli could lead to segregated or fused percepts depending on different instructions given to listeners (Micheyl and Oxenham, 2010b). Therefore, one has to consider the possibility that the role of an acoustic cue on perceptual segregation may be two-fold. First, the acoustic cue may promote pre-attentive saliency of the concurrent sound sources, leading to “primitive” bottom-up segregation (Bregman, 1990, Chap. 4). Second, the acoustic cue may serve as a “clue” that enables listeners to form specific listening strategies, which can then impact perceptual segregation via top-down mechanisms. The current study investigates the concurrent segregation of two simultaneous sounds and the effect of cueing the frequency content of one of the sounds (using a precursor sound) on segregation in an effort to identify the contributions of bottom-up and top-down mechanisms associated with the precursor cue.

Previous research has demonstrated that cueing the frequency of a target tone when the target frequency is randomly drawn from trial to trial benefits its detectability and enhances its perceived pitch (Schlauch and Hafter, 1991; Hafter *et al.*, 1993; Demany *et al.*, 2004; Scharf *et al.*, 2007; Reeves and Scharf, 2010). For example, Scharf *et al.* (2007) measured the detection threshold for 46-ms pure-tone targets in broadband noise under two conditions. In the frequency-certain condition, the target frequency was fixed within a block of trials; while in the frequency-uncertain condition, the target frequency was randomly chosen from nine potential frequencies spanning across a 2.6-octave range. Detection thresholds were about 3–5 dB higher under the frequency-uncertain condition. However, when a 46-ms precursor tone of the same frequency as the target was introduced prior to the presentation of the target, the thresholds in the frequency-uncertain condition improved and approached those in the frequency-certain condition when the precursor-to-target delay (onset to onset) was sufficiently long. The authors argued that the presence of the precursor provided a cue for the listeners to focus their attention to the target frequency (i.e., a top-down process), hence reducing the influence of frequency uncertainty.

Demany *et al.* (2004) studied the effect of cueing one component in a three-tone complex using a precursor tone that was close in frequency to the cued component. The precursor and the tone complex were 700 ms and separated by an inter-stimulus interval (ISI) of 300 ms. An additional comparison tone was presented after the tone complex and the listeners were instructed to identify whether the pitch of the comparison tone was higher or lower than the cued component in the complex. The presence of the precursor benefited the frequency discrimination between the cued component and the comparison tone. Since the frequencies

^{a)}Electronic mail: shen2@indiana.edu

of the tone complex were randomly drawn from trial to trial, the precursor may have helped the listeners orient their attentional focus toward the cued component, leading to enhanced pitch perception. In addition, the benefit was observed not only when the complex and comparison were temporally close (0 or 300 ms), but also when they were 4 s apart, which points to the additional facilitatory role of the precursor in pitch retention.

In some situations, cueing the frequency content of one sound can enhance the audibility and pitch perception of a simultaneous un-cued sound. When a target pure tone is embedded in a simultaneous masking tone-complex, presenting a precursor that consists of the same frequency components as the masker would lead to a “pop-out” percept for the target tone. This is commonly referred to as the auditory enhancement phenomenon (e.g., Viemeister, 1980). Recent studies have suggested that the enhancement phenomenon may consist of both sensory and non-sensory components (e.g., Kidd *et al.*, 2011; Erviti *et al.*, 2011; Cao and Richards, 2012; Carcagno *et al.*, 2012, 2013). The sensory component of the enhancement phenomenon is believed to be related to frequency-specific suppression or inhibition mechanisms along the auditory pathway. When the duration of the precursor is sufficiently long, the suppression/inhibition would be adapted, leading to a stronger neural representation of the later-introduced target than the masker (e.g., see Viemeister and Bacon, 1982; Wright *et al.*, 1993; Byrne *et al.*, 2011; and Shen and Richards, 2012b for psychophysical investigations and Palmer *et al.*, 1995; Nelson and Young, 2010; Carcagno *et al.*, 2014; and Beim *et al.*, 2015 for physiological investigations of the adaptation of suppression/inhibition hypothesis).

The potential existence of a non-sensory component of the enhancement phenomenon has been supported by various psychophysical studies using complex maskers that consisted of randomly drawn frequency components (e.g., Neff and Green, 1987; Richards and Neff, 2004; Richards *et al.*, 2004). In such cases, presenting a precursor that matches the spectrum of the masker can lead to an improvement in the target threshold by as much as 30 dB. This cannot be fully accounted for by the sensory component of the enhancement phenomenon in many situations. A similar effect has also been demonstrated using an experimental paradigm that probes the identification of the target rather than its audibility (Erviti *et al.*, 2011; Carcagno *et al.*, 2012, 2013; Demany *et al.*, 2013; Feng and Oxenham, 2015). Therefore, the hypothesized non-sensory component of the enhancement phenomenon could reflect the attentional focus to the target frequency caused by cueing the frequency content of the masker, which strengthens the perceptual prominence of the target as a segregated auditory object (e.g., Kidd *et al.*, 2011).

Although the effect of frequency cueing on the perceptual segregation of concurrent sounds has not been explicitly studied, many studies investigated the effect of onset asynchrony (i.e., when the precursor is immediately followed by the cued sound) on the segregation of two simultaneous sounds (e.g., Bregman and Pinker, 1978; Summerfield and Assmann, 1989; Hall and Grose, 1991; Darwin and Ciocca, 1992; Moore and Shailer, 1992; Shen and Richards, 2012a).

These studies utilized experimental tasks that encouraged perceptual segregation and inferred the degree of segregation from task performance. For example, Summerfield and Assmann (1989) examined listeners' ability to identify a pair of concurrently presented vowels of the same fundamental frequency. By presenting a 1-s precursor that matched the spectrum of one of the vowels, the accuracy in identifying the un-cued vowel improved significantly and approached that of isolated vowels. This result has been replicated in many other concurrent-vowel studies (e.g., Lentz and Marsh, 2006; Hedrick and Madix, 2009; Valentine and Lentz, 2012). Another common experimental paradigm to investigate the effect of onset asynchrony on perceptual segregation is based on the phenomenon of modulation detection interference (MDI; Yost and Sheft, 1989). The detection or discrimination threshold of amplitude modulation (AM) in a target pure tone is elevated by the concurrent presentation of another tone (interferer) on which AM is imposed. Moreover, the presence of a precursor that matches the frequency of the interferer would lead to a reduction in MDI (Hall and Grose, 1991; Moore and Shailer, 1992; Oxenham and Dau, 2001; Koopman *et al.*, 2008).

While it is evident that in both concurrent-vowel and MDI studies the presence of the precursor (or onset asynchrony) provides a strong acoustic cue for concurrent-sound segregation, it is still unclear whether the segregation originates from solely stimulus-driven, bottom-up processes or if it also involves top-down attentional modulations. Here, we hypothesize that there exist top-down processes that utilize the frequency cue carried by the precursor to facilitate perceptual segregation. To evaluate this hypothesis, the current study adopted an experimental paradigm that addresses perceptual segregation and is designed to differentiate the potential involvements of bottom-up and top-down mechanisms. Following the stimulus paradigm developed by Hall and Grose (1991, Exp. III), two pure tones of equal intensity and different frequencies were simultaneously presented and AM was imposed on either the lower or the higher frequency tone. The listeners were instructed to report which of the two tones, the lower or higher frequency tone, carried the AM by responding “Low” or “High.” The frequencies of the tone pair were randomly drawn on a trial by trial basis so that the listeners were not able to perform the task by discriminating stimuli across trials, but were forced to compare the envelopes of the two simultaneous tones within the single interval on every trial. This is only possible when the two tones are perceived as segregated auditory objects. That is, the performance of this task is not limited by the sensitivity to AM, but limited by the ability to identify whether the AM tone is the lower or higher frequency tone. In addition, it was expected that presenting a precursor tone cueing the lower of the two frequencies would improve the task performance compared to the no-precursor condition. According to our hypothesis, the precursor should allow for attentional focus to either cued or un-cued frequencies, which promotes the perceptual segregation of the two tones.

Several considerations were made to differentiate the bottom-up phenomena such as the sensory component of the auditory enhancement from the hypothesized top-down

contribution to the precursor effect. First, to prevent peripheral interactions between the pair of simultaneous tones, the frequencies of the tone-pair were more than one octave apart hence peripheral suppression was not expected (e.g., Shannon, 1976). Second, the ISI from the precursor to the tone-pair was systematically varied. If bottom-up processes dominate the precursor effect, then the effect should decrease as the ISI increases (e.g., Feng and Oxenham, 2015). Third, multiple precursor durations were tested. Long precursor durations should lead to stronger bottom-up segregation while top-down segregation may not depend on precursor duration since longer precursors do not necessarily carry more information with regard to the frequencies of the tone-pair than shorter precursors. Last, the current study included experiments in which the precursor was presented to either the ipsilateral (Exp. I) or the contralateral (Exp. II) ear. Although bottom-up mechanisms are not limited to the ipsilateral ear (e.g., Carcagno *et al.*, 2012), their influences are expected to be significantly reduced for contralateral precursors (e.g., Viemeister, 1980; Carlyon, 1989; Serman *et al.*, 2008). By systematically varying the ISI, precursor duration, and precursor ear, support for the involvement of top-down (in particular, attentional) processes would be obtained if the observed precursor effect contradicts the expectations from bottom-up mechanisms.

II. EXPERIMENT I: THE EFFECT OF PRECURSOR DURATION AND TEMPORAL SEPARATION

A. Methods

1. Listeners

Eight listeners (four females) were recruited from the student population at Indiana University. All listeners had audiometric thresholds equal to or better than 15 dB hearing level between 250 and 4000 Hz in both ears. For each listener, the ear with the lower pure tone average (PTA) threshold (mean of the hearing thresholds at 0.5, 1, and 2 kHz) was tested. If the PTA thresholds were the same for both ears, the left ear was tested. All of the listeners had practiced for more than 4 h on an experiment similar to the present experiment. Three of the listeners had previously participated in other psychoacoustic experiments.

2. Stimuli

The stimuli were two simultaneously presented tones that differed in frequency (see Fig. 1 for a schematic of the stimulus). These two tones will be referred to as the low tone and high tone based on whether they had a lower or higher frequency. The frequency of the low tone was drawn from a uniform distribution spanning from one-half octave below to one-half octave above 1 kHz on a logarithmic frequency scale. The frequency of the high tone was 1.1 octaves above that of the low tone. Either the low or high tone was amplitude-modulated, determined at random for each trial, forming Low trials (left panel of Fig. 1) or High trials (right panel of Fig. 1), respectively. The modulation rate was fixed at 10 Hz and the modulation depth was $-10 \text{ dB } [20 \log(m)]$. The initial phase of the sinusoidal modulator was randomly

FIG. 1. A schematic of the stimuli for the current experiments. On the High trials (left), two simultaneous pure tones were presented with the higher frequency tone (high tone) being amplitude modulated. The AM is represented in the figure by the presence of two side frequency components on either side of the frequency for the high tone. A precursor tone, when present, was presented at the frequency of the lower frequency tone (low tone). The stimuli on the Low trials (right) were identical, except that the AM was imposed on the low tone. In each panel, T , T_{pre} , and ISI indicate the duration of the simultaneous tone-pair, the duration of the precursor, and the ISI between the precursor and tone-pair, respectively.

drawn from a uniform distribution spanning from 0 to 2π . In most of the experimental conditions a precursor was presented before the low and high tones. The precursor frequency matched that of the low tone. The duration of the precursor (T_{pre}) was 100, 200, or 400 ms in separate conditions. In addition, the ISI between the offset of the precursor and the onset of the low tone was systematically varied from 0 to 1 s (0, 50, 100, 200, 500, and 1000 ms). The high tone, low tone, and precursor were gated on and off using 5-ms raised-cosine ramps and were presented at 70 dB sound pressure level. Note that the level changes caused by imposing AM were accounted for so the low and high tones were always equal in intensity.

The stimuli were generated digitally and presented at a sampling rate of 22.05 kHz using custom software running in MATLAB (The MathWorks, Inc., Natick, MA), which was also used for experimental control and the presentation of visual feedback on correctness following each trial. The stimuli were presented monaurally in the test ear through a 24-bit sound card (MOTU Microbook II, Mark of the Unicorn, Inc., Cambridge, MA) and headphones (HD-280 Pro, Sennheiser Electronic GmbH & Co. KG, Wedemark, Germany). During the present experiment, listeners were seated in a sound-attenuating booth.

3. Procedure

During the present experiment, in a single-interval, 2-alternative forced-choice task, the listeners identified whether the low tone or the high tone was amplitude-modulated by responding Low or High, respectively. Listeners received visual feedback with regard to the response correctness following each response. Due to the frequency randomization, the listeners were not able to discriminate stimuli across trials; rather they had to compare the pitch and fluctuation strength of the two tones on each trial. If the low tone and high tone were perceived as an integrated auditory object instead of two segregated objects, the listeners would not be able to perform the task (Hall and Grose, 1991). Moreover, a relatively high modulation depth was chosen to ensure that the AM was readily detectable and that its detectability would not dominate the performance (see also Sec. IV C). Performance thresholds were determined using a two-down, one-up adaptive procedure, which is associated with the 70.7% position on the psychometric function. To estimate the

performance threshold, the duration of the low and high tones (T) was adaptively changed (indicated by the horizontal double arrows in Fig. 1). The initial duration was 3.16 s. The duration was reduced by a factor of 3.16 following two consecutive correct responses and increased by the same factor following a single incorrect response. The stepping factor was reduced to 1.78 after the first two reversals of the adaptive track, and was further reduced to 1.58 after the fourth reversal. Each adaptive track terminated after 12 reversals were reached. The threshold estimation was based on the geometric mean of the durations at the last eight reversals.

The duration-based adaptive procedure was utilized because it was expected that longer durations would facilitate the segregation of the two tones due to the buildup phenomenon often observed for auditory streaming experiment (e.g., Bregman, 1978; Anstis and Saida, 1985; Roberts *et al.*, 2008; Thompson *et al.*, 2011). A pilot study measuring the psychometric function relating the duration T to the task performance verified this expectation and contributed to the selection of the stepping factors. Threshold estimation based on duration was advantageous over either intensity or AM depth because increasing the intensity or AM depth of one of the tones relative to the other is likely to introduce additional cues that cause the two tones to be perceptually more distinct. Since the current focus is the effect of frequency cueing on segregation through the precursor, these additional segregation cues would complicate the interpretation of the results. One additional advantage of the duration-based adaptive track was that it was able to accommodate for very large individual differences. The duration required to reach threshold could be 10 times longer for some listeners than for others. If a fixed duration was used and the performance was quantified in terms of percent correct, many listeners would reach the performance floor or ceiling.

For each listener, the four precursor conditions (no precursor, $T_{\text{pre}} = 100, 200$, and 400 ms) were tested in a random order. When the precursor was present, the six ISI conditions were tested in a random sequence. Once a threshold was estimated for each of the experimental conditions, the procedure described above was repeated three more times with independent random sequences. The reported thresholds were based on the geometric mean across the four repetitions.

B. Results

In Fig. 2, the individual performance thresholds are plotted as a function of ISI for precursor durations of 100 (top-left), 200 (top-right), and 400 (bottom) ms. Large individual differences were observed across listeners. In certain conditions, the threshold difference was as large as a factor of 10 from the best performer to the worst performer. Despite these large individual differences, certain consistent patterns were observed. For each listener, the threshold duration was generally lower when the precursor was present. In addition, for the precursor durations of 100 and 200 ms (the top panels of Fig. 2), the best performance was not always found when the ISI was 0 ms. For many listeners, the lowest thresholds were found for the longest ISI.

FIG. 2. Individual performance thresholds from Exp. I as a function of ISI for T_{pre} 's of 100, 200, and 400 ms (separate panels). The left-most data points within each panel indicate thresholds obtained without the precursor (the label "NP" is used to indicate the no precursor condition).

Figure 3 plots the precursor effect in terms of the ratio of the threshold obtained without the precursor to the threshold obtained with the precursor. The threshold ratio is shown as a function of ISI for the 100-, 200-, and 400-ms precursor durations (as separate curves). The presence of the precursor improved the threshold for all ISI and T_{pre} conditions tested. When the ISI equaled 0 ms, the threshold ratio was greater for the 400-ms precursor duration than the 100- and 200-ms durations, which may reflect that the longer precursor durations were associated with longer onset asynchrony. For long ISI's (≥ 200 ms), the threshold ratio was comparable across the three values of T_{pre} tested. A repeated measures analysis of variance (ANOVA) was conducted, treating T_{pre} and ISI as the independent variables and (logarithmically transformed) threshold ratio as the dependent variable. A significant main

FIG. 3. Average precursor effect, i.e., the average threshold improvements (in terms of ratio) due to the presence of the precursor, from Exp. I as a function of ISI for T_{pre} 's of 100, 200, and 400 ms (different symbols). A threshold ratio of unity (dashed horizontal line) indicates no precursor effect. The geometric means across the threshold ratios for the eight listeners are shown. Error bars indicate the 68% confidence interval for the mean.

effect of ISI was found [$F(5, 35) = 3.827, p = 0.007$], while the effect of T_{pre} did not reach statistical significance [$F(2, 14) = 2.133, p = 0.155$]. In addition, a significant interaction between T_{pre} and ISI was found [$F(10, 70) = 2.029, p = 0.004$].

The observations that the threshold tended to increase with increasing ISI for the 100- and 200-ms precursors and that the precursor's facilitatory effect on concurrent-sound segregation was evident for ISI's as long as 1 s cannot be explained by stimulus-driven, bottom-up processes alone, such as those underlying the sensory component of the auditory enhancement phenomenon. If the precursor effect observed in the current experiment is associated with top-down central mechanisms, then it would be expected that similar effects will also be found when the precursor is presented to the contralateral ear. This was examined in Exp. II.

III. EXPERIMENT II: CONTRALATERAL PRECURSORS

A. Methods

Five listeners (three females) were included in the current experiment. All had also previously participated in Exp. I. The stimuli and procedure for Exp. II were identical to those of Exp. I, except that the precursor was presented in the contralateral ear and the precursor duration was 100 or 200 ms. Because the no-precursor condition was tested in both Exps. I and II, the reported threshold improvement for each individual listener (e.g., data plotted in Fig. 5) was referenced to the geometric mean of the thresholds in the no-precursor condition across the two experiments for the same listener.

B. Results

Figure 4 plots the individual performance thresholds as a function of ISI for the 100-ms (left panel) and 200-ms (right panel) precursors. Similar to Exp. I, large individual differences were observed. For the condition with no precursor, a threshold difference of up to a factor of 6 was observed from the best performer to the worst performer. Threshold improvement was not observed in all subjects. However, the listener who showed the most improvement from the presence of the precursor sustained the benefit up to 1 s. One listener exhibited a consistent increase in threshold with the presence of the precursor; the same listener (upward triangles in Fig. 4) also had the highest threshold across all

FIG. 4. Individual performance thresholds from Exp. II as a function of ISI for T_{pre} 's of 100 and 200 ms (separate panels). The left-most data points within each panel indicate thresholds obtained without the precursor (NP).

conditions tested compared to the other listeners. The left panel of Fig. 5 summarizes the precursor effect in terms of the ratio of the threshold obtained without the precursor to the threshold obtained with the presence of the precursor as a function of ISI for the 100-ms (circles) and 200-ms (triangles) precursors. The geometric mean of the threshold ratios across listeners is shown in the figure; the geometric mean of the threshold ratio for the same five listeners from Exp. I is shown in the right panel of Fig. 5 for comparison.

In general, the precursor effect, i.e., the threshold ratio, from Exps. I and II were very similar. A repeated measures ANOVA treating precursor ear (ipsilateral versus contralateral), T_{pre} , and ISI as the three independent variables and logarithmically transformed threshold ratio (refer to Fig. 5) as the dependent variable showed a significant main effect of ISI [$F(5, 20) = 9.030, p < 0.001$] and a significant interaction between ISI and T_{pre} [$F(5, 20) = 6.217, p = 0.001$]. The effect of T_{pre} was not significant [$F(1, 4) = 3.093, p = 0.153$]. There was no significant effect of precursor ear [$F(1, 4) = 0.129, p = 0.738$], nor were there any significant interactions involving precursor ear [precursor ear $\times T_{\text{pre}}$: $F(1, 4) = 4.612, p = 0.098$; precursor ear $\times$ ISI: $F(5, 20) = 1.677, p = 0.186$; and precursor ear $\times T_{\text{pre}} \times$ ISI: $F(5, 20) = 1.199, p = 0.345$]. Results indicate that the precursor effect on the segregation of a pair of pure tones is not sensitive to whether the precursor is presented to the ipsilateral or contralateral ear relative to the tone-pair. *Post hoc* pairwise comparisons (Bonferroni corrected) showed that the threshold ratio at the 200-ms ISI was significantly greater than those at ISI's of 0 [$t(4) = 11.045, p = 0.006$], 50 [$t(4) = 16.027, p = 0.001$], and 100 [$t(4) = 11.962, p = 0.004$] ms. Moreover, the threshold ratio at the 100-ms ISI was also greater than that for the 50-ms ISI [$t(4) = 9.438, p = 0.011$]. Therefore, the precursor effect increases with ISI up to 200 ms for T_{pre} 's of 100 and 200 ms.

IV. GENERAL DISCUSSION

A. Evidence for top-down involvements

The current experiments demonstrated that the presence of the precursor facilitated the perceptual segregation of the

FIG. 5. Average precursor effect, in terms of the threshold ratio similar to Fig. 3, from Exps. I and II as functions ISI for T_{pre} 's of 100 and 200 ms (different symbols) and for contralateral (left panel) and ipsilateral (right panel) precursors. A threshold ratio of unity (dashed horizontal line) indicates no precursor effect. The geometric means across the threshold ratios for the five listeners who participated in both Exps. I and II are shown. Error bars indicate the 68% confidence interval for the mean.

low and high tones. However, the effect of ISI on the precursor effect was not consistent with bottom-up segregation mechanisms, which would predict a greater precursor effect when the precursor and tone-pair are at closer temporal proximity (e.g., [Feng and Oxenham, 2015](#)). In contrast, a component of the precursor effect persisted for ISI's as long as 1 s and was not significantly affected as the precursor duration was varied. This observation supports the hypothesis that top-down processes are involved in the precursor's facilitatory effect on concurrent-sound segregation.

[Carlyon \(1989, Exp. 4\)](#) measured the detection of a 1000-Hz, 5-ms tonal target in a simultaneous masker that consisted of two noise bands above and below the target frequency (i.e., notched-noise masker). When a 200-ms precursor with the same spectrum as the masker was introduced, the target threshold was reduced, demonstrating an enhancement phenomenon. Moreover, as the ISI between the precursor and the masker increased, the precursor effect decreased, and became negligible when the ISI exceeded 320 ms. [Cao and Richards \(2012\)](#) studied the effect of a 150-ms precursor as a function of ISI in an enhancement experiment using random-frequency maskers. Their results showed a rapid decay of the precursor effect at very short ISI's followed by a much more gradual decay for long ISI's. The enhancement remained observable as the ISI increased to 4 s. The authors hypothesized that the fast and slow decays are associated with the sensory and non-sensory components of the enhancement phenomenon, respectively. Using a slightly different experimental paradigm, [Carcagno et al. \(2012\)](#) measured listeners' ability to identify a frequency component in a tone complex that was enhanced by presenting a 500-ms precursor sound. As the ISI between the precursor and tone complex increased from 10 to 600 ms, the observed precursor effect decreased exponentially. Contrary to these previous observations, the current study showed that the greatest precursor effect on concurrent sound segregation may not be found at an ISI of 0 ms. For the two shorter precursor durations tested (100 and 200 ms), the precursor effect tended to increase as the ISI increased from 0 to 200 ms.

On the other hand, the currently observed dependency of the precursor effect on ISI is in line with those observed in experiments in which the detection of a random-frequency pure tone in noise was measured (e.g., [Scharf et al., 2007](#); [Reeves and Scharf, 2010](#)). In these studies, the detection threshold of a 46-ms target tone in broadband noise with the target frequency randomly drawn from a 2.6-octave range was measured. A tonal precursor cueing the frequency of the target was presented at various ISI's preceding the target. [Reeves and Scharf \(2010\)](#) found that the detection threshold improved gradually as the ISI increased from 6 to 306 ms. These authors hypothesized that the threshold improvement over increasing ISI could be a combined effect of top-down frequency cueing and frequency-specific interference from the precursor to the target that decreased with increasing ISI. The interference effect proposed by [Reeves and Scharf \(2010\)](#) may not explain the current results. This is because the current experiment was based on a task that involved AM processing rather than the detection of tones. Preceding an AM stimulus with an unmodulated precursor often leads

to improved, rather than degraded, AM detection ([Sheft and Yost, 1990](#); [Bacon and Grantham, 1992](#)). Therefore, low-level interference is not expected to play a significant role in the current study, and the dependency of the precursor effect on ISI observed here is likely to reflect the time course of the top-down processes associated with frequency cueing.

B. The effect of ipsilateral versus contralateral precursors

Experiment II demonstrated that the precursor effect on the segregation of a pair of pure tones was comparable whether the precursor was presented to the ipsilateral or contralateral ear relative to the tone-pair. When compared to previously reported results from the studies of the auditory enhancement phenomenon, the current data were more similar to those obtained from the enhancement experiments with spectral uncertainty ([Richards et al., 2004](#); [Kidd et al., 2011](#); [Erviti et al., 2011](#); [Carcagno et al., 2012](#)) than those obtained using spectrally fixed stimuli ([Viemeister, 1980](#); [Carlyon, 1989](#); [Serman et al., 2008](#)).

[Richards et al. \(2004\)](#) measured listeners' ability to identify the presence of a 1-kHz pure-tone target in a six-tone complex masker. The masker components were randomly drawn from trial to trial, but were distanced from the target frequency by at least 12%. The target was present for 50% of the trials, and each trial was preceded by either the masker alone (M precursor) or a masker-target mixture (SM precursor) at equal probability. In this experiment, the precursor duration was 100 ms and the ISI between the precursor and the masker was 350 ms. The enhancement effect was revealed as a performance advantage associated with the M precursor over the SM precursor. Presenting the precursor to the contralateral ear undermined task performance, but it did not eliminate the enhancement effect. [Kidd et al. \(2011\)](#) asked their listeners to identify narrowband four-tone sequences embedded in spectrally flanking masker sequences. When the frequency region for the target tone-sequence was randomly determined from trial to trial, a preview of the masker sequences presented 250 ms before the trial led to an improvement in performance, demonstrating an enhancement effect. A contralateral precursor was found to be sufficient in eliciting the effect. [Erviti et al. \(2011\)](#) measured listeners' ability to discriminate a target component in a random-frequency tone complex and a pure-tone probe. Since the task required explicit identification of the target component, enhancing the target component by introducing a precursor improved the task performance with similar effects observed for ipsilateral and contralateral precursors.

The dependency of the precursor effect on the precursor ear found in Exp. II resembles that in the enhancement experiment, which has been hypothesized to be governed mainly by the non-sensory component of the enhancement effect ([Richards et al., 2004](#)). This provides additional evidence that the attention focusing process, which is believed to be associated with frequency cueing and the hypothesized non-sensory component of the enhancement phenomenon, can also lead to improved segregation of concurrent sound sources. However, it is worth pointing out that the use of

contralateral precursors cannot entirely exclude bottom-up mechanisms (Carcagno *et al.*, 2012). The lack of effect from the precursor ear alone does not provide conclusive support for the top-down involvement. Rather, the effects of both the precursor ear and ISI have to be considered together. Because the precursor effect increased with ISI for both ipsilateral and contralateral precursors for ISI's from 0 to 200 ms, it is likely that the mechanisms that governed the precursor effect was the same for both precursor-ear conditions, and they involved top-down processes.

C. Considerations on AM sensitivity

The conclusion that top-down processes contribute to concurrent-sound segregation were based on the assumption that the experimental task adopted here reflected perceptual segregation rather than merely the sensitivity to AM. However, it is known that AM sensitivity depends on many temporal factors (e.g., duration, carrier gating, etc.). Because some of these factors were affected in the current experiments as the ISI, T_{pre} , and T were varied, it is conceivable that the observed precursor effect only reflected the precursor's effect on AM sensitivity rather than segregation. Throughout the two experiments, all reported threshold durations (Figs. 2 and 4) were longer than 100 ms. Therefore, it is important to verify whether a 100-ms AM stimulus modulated at 10 Hz and at a depth of -10 dB is expected to be above the AM detection threshold.

When the modulator and carrier of an amplitude-modulated stimulus are gated on and off together (i.e., a gated AM stimulus), the total duration of the stimulus influences the detection or discrimination of the AM (e.g., Sheft and Yost, 1990; Lee and Bacon, 1997; Edwards *et al.*, 2008). A higher modulation depth is required for the listener to detect the presence of AM for shorter durations (e.g., Viemeister, 1979; Sheft and Yost, 1990; Dau *et al.*, 1997; O'Connor *et al.*, 2011). For a 100-ms AM stimulus modulated at 10 Hz, the AM detection threshold is between -15 and -20 dB (Sheft and Yost, 1990; Dau *et al.*, 1997). Moreover, when the AM stimulus is preceded with an unmodulated fringe, similar to introducing a precursor with an ISI of 0 ms in the current experiments, the AM detection threshold is expected to be reduced compared to the gated AM stimulus (Sheft and Yost, 1990; Bacon and Grantham, 1992). Therefore, the -10 -dB modulation depth used in the current experiments was above the AM detection threshold expected for isolated AM stimuli.

When an AM target tone is gated on and off with a second unmodulated masker tone at a remote frequency from the target, the AM detection threshold is minimally affected, leading to a threshold increase typically no more than 5 dB (e.g., Yost and Sheft, 1989; Yost and Sheft, 1994; Oxenham and Dau, 2001; Shen and Lentz, 2010). Moreover, when the masker is gated on prior to the onset of the target, the AM detection threshold is not affected by the amount of onset asynchrony (Koopman *et al.*, 2008). Therefore, for two simultaneously presented pure tones, with one amplitude-modulated at a modulation depth of -10 dB and the other

unmodulated, it is expected that the AM detection threshold would be lower than -10 dB.

In summary, though it was not directly measured, the AM detection thresholds were expected to be better than the -10 -dB modulation depth used in the current experiments across all stimulus durations tested and all precursor conditions. This ensured that the results reported from the current study were not dominated by AM sensitivity but instead by perceptual segregation.

V. SUMMARY

The current study investigated whether cueing the frequency of one tone in a simultaneous pair of pure tones (using a precursor tone) affected the perceptual segregation of the two tones and whether this precursor effect had a component that was related to top-down processes. A sole bottom-up mechanism would predict that the precursor effect would increase when (1) the precursor duration is longer, (2) the ISI between the precursor and the tone-pair is shorter, and (3) the precursor is presented to the ipsilateral rather than the contralateral ear relative to the tone-pair. None of the above predictions were consistently observed in the current experiments, indicating the potential involvement of top-down mechanisms.

ACKNOWLEDGMENTS

This work was supported by the NIH Grant No. R21 DC013406 (MPIs: V.M.R. and Y.S.). Brooke Louthan provided assistance for data collection. Carolyn J. Herbert and Allison B. Kern provided assistance in manuscript preparation.

- Anstis, S. M., and Saida, S. (1985). "Adaptation to auditory streaming of frequency-modulated tones," *J. Exp. Psychol.-Hum. Percept. Perform.* **11**, 257–271.
- Bacon, S. P., and Grantham, D. W. (1992). "Fringe effects in modulation masking," *J. Acoust. Soc. Am.* **91**, 3451–3455.
- Beim, J. A., Elliott, M., Oxenham, A. J., and Wojtczak, M. (2015). "Stimulus frequency otoacoustic emissions provide no evidence for the role of efferents in the enhancement effect," *J. Assoc. Res. Otolaryngol.* **16**, 613–629.
- Bregman, A. S. (1978). "Auditory streaming is cumulative," *J. Exp. Psychol.: Human Percept. Perform.* **4**(3), 380–387.
- Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organisation of Sound* (MIT Press, Cambridge, MA), 792 pp.
- Bregman, A. S., and Pinker, S. (1978). "Auditory streaming and the building of timbre," *Can. J. Psychol.* **32**, 19–31.
- 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**, 2088–2094.
- Cao, X., and Richards, V. M. (2012). "Enhancement in informational masking," *J. Speech Lang. Hear. Res.* **55**, 1135–1147.
- Carcagno, S., Plack, C. J., Portron, A., Semal, C., and Demany, L. (2014). "The auditory enhancement effect is not reflected in the 80-Hz auditory steady-state response," *J. Assoc. Res. Otolaryngol.* **15**, 621–630.
- Carcagno, S., Semal, C., and Demany, L. (2012). "Auditory enhancement of increments in spectral amplitude stems from more than one source," *J. Assoc. Res. Otolaryngol.* **13**, 693–702.
- Carcagno, S., Semal, C., and Demany, L. (2013). "No need for templates in the auditory enhancement effect," *PLoS One* **8**, e67874.
- Carlyon, R. P. (1989). "Changes in the masked thresholds of brief tones produced by prior bursts of noise," *Hear. Res.* **41**, 223–235.

- Dannenbring, G. L., and Bregman, A. S. (1978). "Streaming vs. fusion of sinusoidal components of complex tones," *Percept. Psychophys.* **24**, 369–376.
- Darwin, C. J., and Ciocca, V. (1992). "Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component," *J. Acoust. Soc. Am.* **91**, 3381–3390.
- Darwin, C. J., and Hukin, R. W. (1997). "Perceptual segregation of a harmonic from a vowel by interaural time difference and frequency proximity," *J. Acoust. Soc. Am.* **102**, 2316–2324.
- Darwin, C. J., Pattison, H., and Gardner, R. B. (1989). "Vowel quality changes produced by surrounding tone sequences," *Percept. Psychophys.* **45**, 333–342.
- Dau, T., Kollmeier, B., and Kohlrausch, A. (1997). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration," *J. Acoust. Soc. Am.* **102**, 2906–2919.
- Demany, L., Carcagno, S., and Semal, C. (2013). "The perceptual enhancement of tones by frequency shifts," *Hear. Res.* **298**, 10–16.
- Demany, L., Montandon, G., and Semal, C. (2004). "Pitch perception and retention: Two cumulative benefits of selective attention," *Percept. Psychophys.* **66**, 609–617.
- Edwards, D. R., Lee, J., Andrews, J., and Wong, A. (2008). "Contribution of onset/offset information of modulation to amplitude modulation depth discrimination," *J. Acoust. Soc. Am.* **123**, EL111–EL115.
- Erviti, M., Semal, C., and Demany, L. (2011). "Enhancing a tone by shifting its frequency or intensity," *J. Acoust. Soc. Am.* **129**, 3837–3845.
- Feng, L., and Oxenham, A. J. (2015). "New perspectives on the measurement and time course of auditory enhancement," *J. Exp. Psychol.-Hum. Percept. Perform.* **41**, 1696–1708.
- Haft, E. R., Schlauch, R. S., and Tang, J. (1993). "Attending to auditory filters that were not stimulated directly," *J. Acoust. Soc. Am.* **94**, 743–747.
- Hall, J. W., III, and Grose, J. H. (1991). "Some effects of auditory grouping factors on modulation detection interference (MDI)," *J. Acoust. Soc. Am.* **90**, 3028–3035.
- Hedrick, M. S., and Madix, S. G. (2009). "Effect of vowel identity and onset asynchrony on concurrent vowel identification," *J. Speech Lang. Hear. Res.* **52**, 696–705.
- Kidd, G., Jr., Richards, V. M., Streeter, T., Mason, C. R., and Huang, R. (2011). "Contextual effects in the identification of nonspeech auditory patterns," *J. Acoust. Soc. Am.* **130**, 3926–3938.
- Koopman, J., Houtgast, T., and Dreschler, W. A. (2008). "Modulation detection interference for asynchronous presentation of masker and target in listeners with normal and impaired hearing," *J. Speech Lang. Hear. Res.* **51**, 1588–1598.
- Lee, J., and Bacon, S. P. (1997). "Amplitude modulation depth discrimination of a sinusoidal carrier: Effect of stimulus duration," *J. Acoust. Soc. Am.* **101**, 3688–3693.
- Lentz, J. J., and Marsh, S. L. (2006). "The effect of hearing loss on identification of asynchronous double vowels," *J. Speech Lang. Hear. Res.* **49**, 1354–1367.
- Michéyl, C., and Oxenham, A. J. (2010a). "Pitch, harmonicity and concurrent sound segregation: Psychoacoustical and neurophysiological findings," *Hear. Res.* **266**, 36–51.
- Michéyl, C., and Oxenham, A. J. (2010b). "Objective and subjective psychophysical measures of auditory stream integration and segregation," *J. Assoc. Res. Otolaryngol.* **11**, 709–724.
- Moore, B. C., and Shailer, M. J. (1992). "Modulation discrimination interference and auditory grouping," *Philos. Trans. Royal Soc. B* **336**, 339–346.
- 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.
- O'Connor, K. N., Johnson, J. S., Niwa, M., Noriega, N. C., Marshall, E. A., and Sutter, M. L. (2011). "Amplitude modulation detection as a function of modulation frequency and stimulus duration: Comparisons between macaques and humans," *Hear. Res.* **277**, 37–43.
- Oxenham, A. J., and Dau, T. (2001). "Modulation detection interference: Effects of concurrent and sequential streaming," *J. Acoust. Soc. Am.* **110**, 402–408.
- 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.
- Rasch, R. A. (1978). "The perception of simultaneous notes such as in polyphonic music," *Acta Acust. Acust.* **40**, 21–33.
- Reeves, A., and Scharf, B. (2010). "Auditory frequency focusing is very rapid," *J. Acoust. Soc. Am.* **128**, 795–803.
- Richards, V. M., Huang, R., and Kidd, G., Jr. (2004). "Masker-first advantage for cues in 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.
- Roberts, B., Glasberg, B. R., and Moore, B. C. (2008). "Effects of the build-up and resetting of auditory stream segregation on temporal discrimination," *J. Exp. Psychol.-Hum. Percept. Perform.* **34**, 992–1006.
- Scharf, B., Reeves, A., and Suci, J. (2007). "The time required to focus on a cued signal frequency," *J. Acoust. Soc. Am.* **121**, 2149–2157.
- Schlauch, R. S., and Haft, E. R. (1991). "Listening bandwidths and frequency uncertainty in pure-tone signal detection," *J. Acoust. Soc. Am.* **90**, 1332–1339.
- 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.
- Sheft, S., and Yost, W. A. (1990). "Temporal integration in amplitude modulation detection," *J. Acoust. Soc. Am.* **88**, 796–805.
- Shen, Y., and Lentz, J. J. (2010). "Effect of fast-acting compression on modulation detection interference for normal hearing and hearing impaired listeners," *J. Acoust. Soc. Am.* **127**, 3654–3665.
- Shen, Y., and Richards, V. M. (2012a). "Spectral processing of two concurrent harmonic complexes," *J. Acoust. Soc. Am.* **131**, 386–397.
- Shen, Y., and Richards, V. M. (2012b). "Investigating the auditory enhancement phenomenon using behavioral temporal masking patterns," *J. Acoust. Soc. Am.* **132**, 3363–3374.
- Summerfield, Q., and Assmann, P. F. (1989). "Auditory enhancement and the perception of concurrent vowels," *Percept. Psychophys.* **45**, 529–536.
- Thompson, S. K., Carlyon, R. P., and Cusack, R. (2011). "An objective measurement of the build-up of auditory streaming and of its modulation by attention," *J. Exp. Psychol.-Hum. Percept. Perform.* **37**, 1253–1262.
- Valentine, S., and Lentz, J. J. (2012). "The influence of reduced audible bandwidth on asynchronous double-vowel identification," *J. Speech Lang. Hear. Res.* **55**, 1750–1764.
- van Noorden, L. P. A. S. (1975). "Temporal coherence in the perception of tone sequences," Ph.D. dissertation, Institute for Perception Research, Eindhoven, The Netherlands.
- Viemeister, N. F. (1979). "Temporal modulation transfer functions based upon modulation thresholds," *J. Acoust. Soc. Am.* **66**, 1364–1380.
- Viemeister, N. F. (1980). "Adaptation of masking," in *Psychophysical, Physiological and Behavioural Studies in Hearing*, edited by G. van den Brink and F. A. Bilsen (Noordwijkerhout, The Netherlands), pp. 190–199.
- Viemeister, N. F., and Bacon, S. P. (1982). "Forward masking by enhanced components in harmonic complexes," *J. Acoust. Soc. Am.* **71**, 1502–1507.
- Wright, B. A., McFadden, D., and Champlin, C. A. (1993). "Adaptation of suppression as an explanation of enhancement effects," *J. Acoust. Soc. Am.* **94**, 72–82.
- Yost, W. A., and Sheft, S. (1989). "Across-critical-band processing of amplitude-modulated tones," *J. Acoust. Soc. Am.* **85**, 848–857.
- Yost, W. A., and Sheft, S. (1994). "Modulation detection interference: Across-frequency processing and auditory grouping," *Hear. Res.* **79**, 48–58.