Asynchronous glimpsing of speech: Spread of masking and task set-size

Erol J. Ozmeral,[a] Emily Buss, and Joseph W. Hall III
Department of Otolaryngology/Head and Neck Surgery, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599

(Received 14 June 2011; revised 11 April 2012; accepted 11 June 2012)

Howard-Jones and Rosen [(1993). J. Acoust. Soc. Am. **93**, 2915–2922] investigated the ability to integrate glimpses of speech that are separated in time and frequency using a "checkerboard" masker, with asynchronous amplitude modulation (AM) across frequency. Asynchronous glimpsing was demonstrated only for spectrally wide frequency bands. It is possible that the reduced evidence of spectro-temporal integration with narrower bands was due to spread of masking at the periphery. The present study tested this hypothesis with a dichotic condition, in which the even- and odd-numbered bands of the target speech and asynchronous AM masker were presented to opposite ears, minimizing the deleterious effects of masking spread. For closed-set consonant recognition, thresholds were 5.1–8.5 dB better for dichotic than for monotic asynchronous AM conditions. Results were similar for closed-set word recognition, but for open-set word recognition the benefit of dichotic presentation was more modest and level dependent, consistent with the effects of spread of masking being level dependent. There was greater evidence of asynchronous glimpsing in the open-set than closed-set tasks. Presenting stimuli dichotically supported asynchronous glimpsing with narrower frequency bands than previously shown, though the magnitude of glimpsing was reduced for narrower bandwidths even in some dichotic conditions.
[http://dx.doi.org/10.1121/1.4730976]

PACS number(s): 43.71.Rt, 43.66.Dc [LD] Pages: 1152–1164

I. INTRODUCTION

In natural settings, such as a noisy city street or crowded party, there is a combination of interfering sounds that fluctuate in time and frequency depending on their sources. Because most natural masking noises tend to vary in their spectro-temporal structure, listeners are sometimes able to take advantage of the redundancy in speech across time and frequency by attending to regions in the stimulus which have the best signal-to-noise ratios (SNRs; Miller and Licklider, 1950; Dirks and Bower, 1970; Howard-Jones and Rosen, 1993). Often, these natural stimuli are comodulated, with coherent envelopes across frequency (Nelken *et al.*, 1999). It is during the low-amplitude portions of modulated maskers that target speech has the best SNR. Taking advantage of the high SNR at the masker minima, also known as glimpsing (Li and Loizou, 2007; Gnansia *et al.*, 2008) or dip-listening (e.g., Peters *et al.*, 1998), typically leads to improved identification.

In one of the earliest studies on the effects of masker modulation on speech intelligibility, Miller and Licklider (1950) observed that performance is highly dependent on the rate of masker fluctuation. As the rate of modulation decreases below 200 Hz, intelligibility increases until around 10 Hz; however, as modulation rates are lowered below 10 Hz, entire words tend to be masked, and subsequently, performance declines. The optimal rate of modulation has been shown to depend on the type of speech material and the number of possible response alternatives (Buss *et al.*, 2009). In addition to studies that have found modulation rate to be an important parameter (Miller and Licklider, 1950; Buss *et al.*, 2009), the amount of masking release incurred by introducing masker amplitude modulation (AM) is larger for deeper masker modulation depth (Gnansia *et al.*, 2008), and for more intense maskers (Summers and Molis, 2004; George *et al.*, 2006).

Whereas most studies of masker fluctuation have evaluated envelope fluctuations that are coherent across frequency, naturally occurring maskers often contain spectro-temporally complex fluctuations. Howard-Jones and Rosen (1993) tested the hypothesis that masking release associated with masker AM depends on the epochs of improved SNR coinciding across frequency. Their innovative design tested noise maskers that were separated into frequency channels, or bands, which spanned 100 to 10 000 Hz in equal log steps. These bands were then amplitude modulated on and off in a square-wave fashion at a rate of 10 Hz. Howard-Jones and Rosen controlled the phase of AM in neighboring bands, so that modulation was either in-phase (synchronous) or 180 degrees out-of-phase (asynchronous). When the AM was out-of-phase in neighboring bands, the masker resembled a checkerboard when viewed by its time-frequency representation, or spectrogram. The task was consonant identification in a vowel-consonant-vowel (VCV) context, and the masker was a pink noise that had no modulation, synchronous AM, or asynchronous AM, with varying numbers of frequency bands. It was found that synchronous AM noise improved thresholds by 23 dB relative to the unmodulated noise condition; that is, there was a 23-dB masking release when the masker was coherently amplitude modulated. In asynchronous AM conditions,

[a]Author to whom correspondence should be addressed. Electronic mail: eozmeral@unc.edu

there was some masking release when noise was filtered into two or four frequency bands—15.5 dB and 6 dB, respectively—but close to zero unmasking was observed in the 8- or 16-band conditions. Interestingly, it was shown that thresholds for the 2-band asynchronous AM condition were significantly higher (i.e., better) than for conditions in which one band was modulated and the other was left unmodulated. This led to the conclusion that masking release in the asynchronous modulation condition was not based solely on information present in a subset of bands, but instead demonstrated speech integration for signals that were unmasked asynchronously across time and frequency. Although thresholds were lower in the 4-band asynchronous AM condition than the unmodulated condition, thresholds were no better in the asynchronous AM condition than in a control condition where two bands were modulated and the other two were left unmodulated. Howard-Jones and Rosen (1993) therefore concluded that although asynchronous glimpsing occurred in the 2-band condition, it was not evident in the 4-, 8-, and 16-band cases.

It remains unclear why Howard-Jones and Rosen (1993) found no evidence of asynchronous glimpsing with greater than two bands. One possibility is that there is a perceptual limit on the ability to integrate asynchronous speech information when speech is distributed across a large number of frequency bands, but other evidence makes this unlikely (Buss *et al.*, 2004). In a speech identification experiment, Buss and colleagues (2004) determined masked identification thresholds for synchronous and asynchronous AM *speech* filtered into 2, 4, 8, or 16 frequency bands. Speech reception thresholds were determined for the modulated speech presented in a steady-state pink noise. Results of this study showed comparable benefit of synchronous AM and asynchronous AM when the speech itself was modulated, regardless of the number of bands. This result provided evidence for spectro-temporal integration of asynchronous speech information even when there were as many as 16 relatively narrow bands.

This AM speech result—that integration is possible for greater than two or four bands of asynchronously modulated speech—prompts consideration of alternative explanations for the failure of Howard-Jones and Rosen (1993) to find evidence of asynchronous glimpsing in parallel conditions where the noise was asynchronously modulated. One possible explanation for why synchronous AM noise had the largest masking release in the data of Howard-Jones and Rosen is that better performance in the synchronous AM noise is aided by comodulation masking release (CMR; Hall *et al.*, 1984). In short, CMR is the improvement in detection thresholds seen when comodulated off-frequency maskers are added to an on-frequency masked target. While CMR could have played some role in the results of Howard-Jones and Rosen, it is unlikely to account for synchronous/asynchronous AM differences that were on the order of 20 dB; studies have shown CMR to have relatively small contributions to performance with supra-threshold stimuli, including speech (Grose and Hall, 1992; Hall *et al.*, 1997; Kwon, 2002; Buss *et al.*, 2003).

Another possibility is that better performance in the synchronous than asynchronous AM noise conditions may be due to spread of masking associated with the asynchronous AM noise (Howard-Jones and Rosen, 1993). Spread of masking is the phenomenon in which a masker in a neighboring frequency region causes substantial energetic masking of a target stimulus. The amount of masking (in dB) is greatest when the masker is relatively intense (Wegel and Lane, 1924; Moore *et al.*, 1998). In the case of asynchronous AM masking, the advantage of selectively listening to unmasked frequency regions of target speech is likely to be reduced due to the spread of masking from the neighboring frequency regions, in which the masker is in the "on"-phase of AM. That is, when an even- or odd-numbered frequency band is in the "off"-phase of modulation, there is a neighboring odd- or even-numbered band, respectively, above and/or below it, which is "on" and contributing energetic masking. This effect is expected to be more detrimental when the frequency bands are narrow, since any spread can mask a larger proportion of the neighboring unmasked region. Hence, each masker has greater potential to degrade performance via spread of masking when there are large numbers of narrow bands, due to close proximity to neighboring bands.

Since listeners can integrate speech information distributed across a large number of asynchronous speech bands under some conditions (Buss *et al.*, 2004), Howard-Jones and Rosen (1993) may have shown only minimal integration because spread of masking degraded the quality of the available speech. Importantly, Howard-Jones and Rosen presented their stimuli diotically, meaning that all stimuli were presented to both ears symmetrically. Since spread of masking occurs when asynchronous AM maskers are summed together at the periphery, it is expected that the effects of masking spread should be greatly diminished or eliminated if the even- and odd-numbered bands are presented to opposite ears. By separating the bands across the ears, the peaks of modulation will no longer exert spread of masking on the dips of modulation in the neighboring bands, providing the listener a better opportunity to identify the speech.

II. EXPERIMENT 1

The first experiment generally followed the methods of Howard-Jones and Rosen (1993), but included dichotic conditions, in which the even- and odd-numbered stimulus bands were presented to opposite ears, and novel control conditions, in which only even- or odd-number speech bands were presented along with full or partial maskers. Dichotic presentation was chosen because it reduces the effect of masking spread at the periphery, which could underlie the fact that Howard-Jones and Rosen did not find asynchronous glimpsing with greater than two bands. The goal was to determine whether asynchronous glimpsing in the Howard-Jones and Rosen study was limited by spread of masking, and whether the auditory system can indeed integrate asynchronous cues for speech identification across time and frequency with narrower spectral bands than seen before.

A. Methods

1. Listeners

Six native English speaking, young adults with no history of hearing loss or ear problems were recruited from the

Chapel Hill community. All listeners were screened for normal hearing, with a criterion of pure tone thresholds of 20 dB hearing level or better at octave frequencies from 250 to 8000 Hz in both ears (ANSI, 1994).

2. Stimuli

The speech material included 12 intervocalic consonants ([b d f g k m n p s t v z] as in [a*m*a]) spoken by an adult female speaker from this lab. There were five recordings of each stimulus, for a total of 60 recordings. These speech tokens were 528–664 ms in duration, with a mean duration of 608 ms. Recordings were made at a 44100-Hz sampling rate, but they were subsequently up-sampled to 48828 Hz to conform to hardware specifications. Each token was digitally scaled so that all samples had an equal root-mean-square (rms) level. These stimuli were then filtered into 2, 4, 8, or 16 frequency bands using sixth-order Butterworth band-pass filters. Filter bandwidth was equivalent in logarithmic units, with bands spanning 100 to 10 000 Hz.

The maskers were pink noise samples that, by definition, contained equal energy per octave band. Each masker sample was generated digitally with duration equal to the longest possible speech token plus 300 ms (964 ms total duration). When speech tokens were present, presentations began 150 ms after the onset of the masker. Modulated maskers were either modulated synchronously (Sync) or asynchronously (Async) across frequency, with a modulation rate of 10 Hz and random starting phase. The following steps were performed to create these stimuli. First, pink noise was filtered using the same procedure and parameters discussed above for the speech stimuli. Second, each frequency band was modulated on and off at 10 Hz, with a starting phase alternating between starting on and starting off for consecutive bands in Async conditions. In order to limit spectral energy to the specified frequency region, 10-ms raised cosines were used to smooth these modulation transitions.

Maskers could be presented either monotically to the left or right ear (L or R, respectively) or dichotically (Δ). Dichotic stimulation presents the odd-numbered bands to the left ear and the even-numbered bands to the right ear. Monotic stimulation was chosen over diotic to avoid diotic summation which can account for nearly 20% better speech reception thresholds (SRTs) than monotic presentations (Davis and Haggard, 1982).[1] When speech bands were present, they were summed with the associated masker bands. In some cases, masker bands were presented without the associated speech bands.

3. Procedure and conditions

An adaptive "up-down" procedure was used to determine the SRTs corresponding to 50% correct identification (Levitt, 1971). The adaptive computer-controlled test procedure used a custom graphical user interface administered through MATLAB on a personal computer. Stimuli were presented through a pair of insert headphones (Etymotic ER-2, Elk Grove Village, IL), and listeners were seated in a single-wall, sound-treated booth. The level of the speech was fixed at 45 dB sound pressure level (SPL) before filtering into bands, and no adjustment of the speech level was made to offset the overall energy reduction due to filtering. The initial masker level was set to 10 dB below pilot threshold levels determined for each condition. The level of the masking noise was turned up or down by 4 dB, depending on whether the previous response was correct or incorrect, respectively. The listener's estimated threshold was determined by computing the mean masker level at the last 24 of 26 track reversals. Thresholds were blocked by condition, and the order of conditions was quasi-randomly selected for each listener to avoid order effects. Each listener performed between three and four tests for each condition. The fourth estimate was obtained if the first three thresholds were not all within 3 dB of each other. Across subjects, this occurred for 14–18 of the 21 conditions. Overall testing time was roughly 4 h, typically spread out over three non-consecutive sessions.

During the test, the speech token associated with each interval was randomly selected with replacement. Listeners responded by clicking a button on the computer screen corresponding to the consonant heard, out of a possible 12 consonants. In all, there were 21 test conditions, which are illustrated in Fig. 1 (see also Table I). All thresholds were referenced to the unmodulated noise condition (Unmod). Two conditions used synchronous AM, one monotic (Sync-R) and one dichotic (Sync-Δ); the Sync-Δ was generated only for the 8-band condition. For each asynchronous monotic and dichotic condition (Async-R and Async-Δ, respectively), stimuli were processed into 2, 4, 8, or 16 bands for a total of eight asynchronous test conditions. The key distinction between monotic (L or R) and dichotic (Δ) configurations is that the former has stimulus bands presented to one ear, whereas the latter has just the even bands presented to the right ear and just the odd bands presented to the left ear.

There were two types of control condition for the Async-Δ conditions. One was just like Async-Δ, but the *speech* bands were present in only one of the ears: in Async-Δ-EVEN, the even speech bands were presented to the right ear, and the even and odd noise bands were presented to the right and left ears, respectively; in Async-Δ-ODD, the odd speech bands were presented to the left ear, and the even and odd noise bands were again presented to the right and left ears, respectively (see Fig. 1). These control conditions were intended to reveal whether performance in the Async-Δ conditions could be accounted for solely by either the even or odd speech bands. These controls were run for all four of the Async band number conditions. Note that in these control conditions, one of the ears receives no speech signal, but does receive masking bands that are "on" when the speech bands in the other ear are unmasked (i.e., the ipsilateral maskers are in "off"-phase). Two additional control conditions were run to determine whether these noise bands, contralateral to the speech bands, had any masking effect. One was like Async-Δ, except only the right ear received input (Async-R-EVEN). The other was like Async-Δ, except that only the left ear received input (Async-L-ODD). This type of control was run only for the 8-band case.

Both types of control conditions used here differed from those used by Howard-Jones and Rosen (1993). In that study, the modulated even (or odd) masker bands were presented

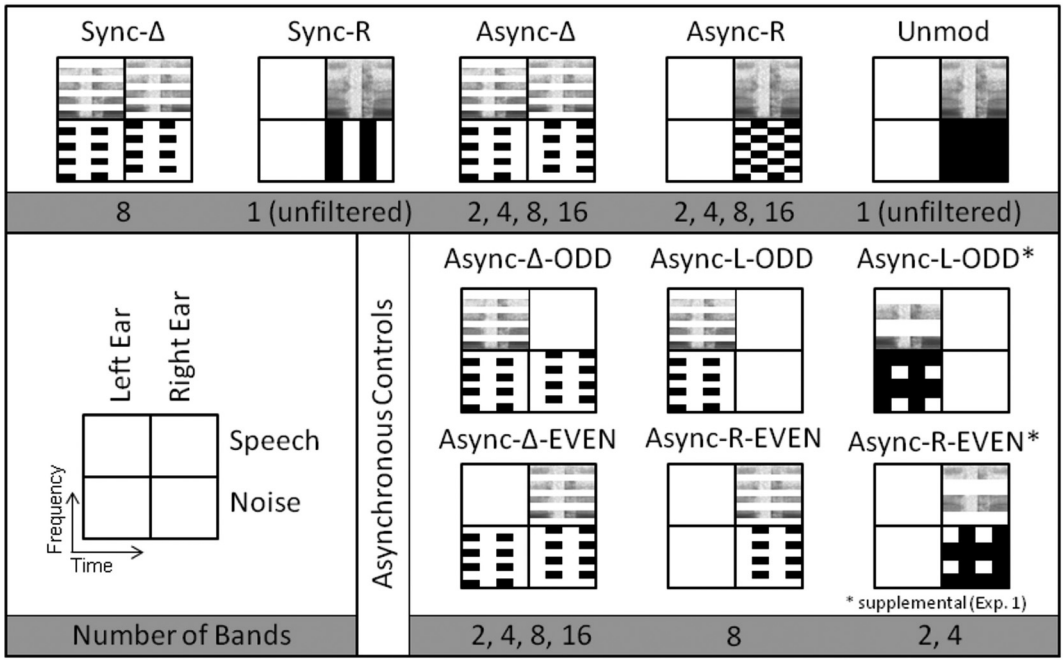

FIG. 1. Schematic of masker conditions in all experiments. Primary conditions are represented on the top row, and controls are shown below. The order of the primary conditions in the top row is an indication of the expected ranking in thresholds, with the best performance starting on the left, with the two Sync conditions, and the worst performance on the right, with the Unmod condition. As the legend indicates, each condition is represented as a 2-by-2 box in which the left and right columns represent stimulation of the left and right ears, respectively, and the top and bottom rows represent the speech and noise stimuli, respectively. In each box, frequency from 0.1 to 10 kHz is represented vertically, and a time span of 200 ms is represented horizontally. Speech is represented via spectrogram, and noise is represented by black spectro-temporal regions indicating the "on" periods of masker modulation. Amplitude modulation was performed at a rate of 10 Hz, and frequency bands were filtered in equal widths on a logarithmic scale. The numbers of bands tested per condition are given in the shaded regions below the conditions.

with an unmodulated masker in the frequency regions associated with the odd (or even) bands. These steady masker bands could introduce spread of masking, similar to that hypothesized in the Async-R conditions. For the purpose of demonstrating spectro-temporal integration of speech in even and odd bands in the Async-Δ conditions, control conditions free from spread of masking were necessary. The present control conditions therefore omit this steady masker in the complementary (non-speech) spectral region.

In this paradigm, masking release is quantified as the difference in threshold between a condition with modulated noise and the unmodulated noise (Unmod) condition. Greater

TABLE I. Mean SRTs (in dB SNR) from experiment 1 are reported for each stimulus condition. The standard error of the mean ($n = 6$) is shown in parentheses below the associated mean. For the dichotic control conditions, the condition associated with better performance is indicated by an asterisk for each number of bands. Recall that control conditions included only half of the speech bands of the associated Async condition; for example, the 8-band controls included only four bands of speech.

Conditions		Unfiltered	Number of bands			
			2	4	8	16
Primary data	Unmod	−1.9 (0.5)				
	Sync-R	−26.7 (2.1)				
	Sync-Δ				−24.2 (1.6)	
	Async-R		−19.1 (0.9)	−15.4 (0.6)	−9.1 (0.6)	−7.8 (0.4)
	Async-Δ		−24.2 (1.5)	−22.8 (1.1)	−17.4 (1.7)	−16.3 (1.6)
Controls	Async-Δ-ODD		−2.6 (1.7)	−6.4 (2.3)	−5.3 (1.4)	−9.1 (1.0)*
	Async-Δ-EVEN		−20.0 (1.6)*	−10.1 (3.0)*	−8.6 (2.2)*	−6.3 (1.3)
	Async-L-ODD				−8.2 (2.5)	
	Async-R-EVEN				−14.4 (1.8)	
Supplemental data	Unmod	−4.7 (0.4)				
	Async-L-ODD + steady EVEN		−7.1 (0.5)	−7.8 (0.3)		
	Async-R-EVEN + steady ODD		−12.5 (2.2)*	−8.5 (1.3)*		

masking release is expected in the Async-Δ than the Async-R conditions, and this difference will be referred to as a "dichotic advantage." The ability to combine information that is separated in time and frequency, "asynchronous glimpsing," is defined as the difference between an Async-Δ condition and the better of the two complementary dichotic control conditions (either Async-Δ-ODD or Async-Δ-EVEN).

B. Results

Figure 2 shows the mean SRTs (in dB SNR) for the asynchronous and synchronous masker conditions as well as the better Async-Δ control, expressed relative to the SRT for unmodulated pink noise. Error bars show one standard error of the mean, and symbols indicate the AM masker conditions, as defined in the legend. Mean SRTs are also presented in Table I for all conditions. The mean SRT in the reference (Unmod) condition is –1.9 dB SNR, and the SRTs for all conditions and bands shown in Fig. 1 are significantly lower than this reference (paired *t*-test; $p < 0.05$). Thresholds in the Sync-R and Sync-Δ conditions are not significantly different ($t_5 = 1.18$, $p = 0.29$), so the average of these two conditions is shown in the data figure (Sync, average).

Figure 2 shows that release from masking (i.e., the absolute difference between a condition and the reference condition) is greatest for the Sync-R and Sync-Δ conditions (average of 23.9 dB better threshold), intermediate for the Async-Δ conditions (ranging from 22.2 to 14.4 dB as band number increases), and least for the Async-R conditions (ranging from 17.1 to 5.9 dB as band number increases). A two-way repeated-measures analysis of variance (ANOVA) was performed to compare performance in the Async-Δ and Async-R conditions, with two levels of condition and four levels of band number. This analysis yielded a main effect of condition ($F_{1,5} = 49.5$, $p = 0.001$), a main effect of the number of bands ($F_{3,15} = 54.4$, $p < 0.001$), and no interaction ($F_{3,15} = 2.20$, $p = 0.13$). The difference in masking release between the dichotic and monotic asynchronous conditions is between 5.1 and 8.5 dB, with greater masking release for the dichotic conditions. If the effect of spread of masking is greater for larger numbers of bands, and if the benefit of dichotic presentation is predominantly due to reduced effects of spread of masking, then the difference between Async-Δ and Async-R conditions should increase with numbers of bands. A planned linear contrast on the condition-by-band interaction indicates a non-significant trend in this direction ($F_{1,5} = 5.51$, $p = 0.06$). It is important to note that the roughly 23-dB release from masking observed in the Sync conditions is the same as that found by Howard-Jones and Rosen (1993). However, in contrast to Howard-Jones and Rosen, this study *does* find that performance for the Async-R condition is better than performance for unmodulated noise at all numbers of bands, though there is a similar reduction in performance as the number of bands increases. A linear contrast in a one-way ANOVA with four levels of number of bands confirmed the increase in thresholds with the number of bands ($F_{1,5} = 201.9$, $p < 0.001$).

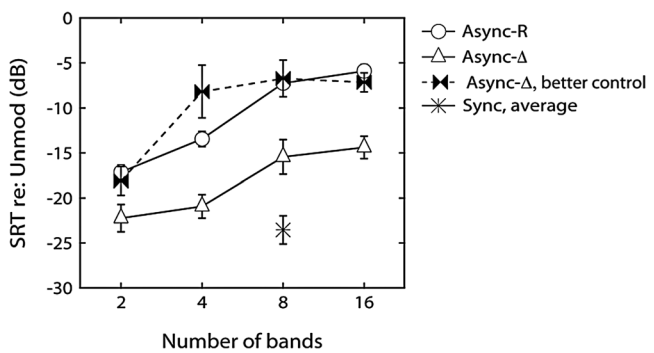

FIG. 2. Mean SRTs in experiment 1 are plotted for modulated noise conditions relative to the unmodulated condition. The difference in mean thresholds relative to the Unmod condition at 2, 4, 8, or 16 bands are plotted for the monotic asynchronous condition (circles), the dichotic asynchronous condition (triangles), the better dichotic control condition (bowties), and the mean of the monotic and dichotic synchronous conditions (8-band only; stars). Error bars indicate standard error of the mean ($n = 6$).

Control measures taken in the study are useful in assessing the possibility that a listener was simply attending to a subset of bands—either the even or the odd bands—for the Async conditions, thereby not actually integrating across frequency *and* time. As can be seen in Fig. 2, performance in the Async-Δ conditions was uniformly superior to that obtained in the Async-Δ-ODD and Async-Δ-EVEN control conditions (Async-Δ, better control). The masking release is 4.1–21.6 dB greater in the Async-Δ conditions than in the dichotic control conditions, depending on the number of bands and the particular subset. To evaluate this statistically, mean SNRs across listeners were compared in the two dichotic control conditions (Async-Δ-EVEN and Async-Δ-ODD). The "better control" condition, the odd or even condition with the better threshold (lower SNR), was identified for each number of bands. Individual data for these better control conditions were evaluated relative to the Async-Δ condition with a repeated-measures ANOVA, including two levels of dichotic condition (Async-Δ and the better control) and four levels of band number (2, 4, 8, and 16). The analysis indicates significant main effects of condition ($F_{1,5} = 45.7$; $p = 0.001$) and the number of bands ($F_{3,15} = 52.0$; $p < 0.001$), and no interaction ($F_{3,15} = 2.7$; $p = 0.08$). This indicates that performance in the Async-Δ condition was consistently better than that possible with either the odd or the even speech bands alone. In other words, listeners were making use of information from spectral regions associated with both the even and odd bands, and this integration was not dependent on the number of bands. This is in contrast to the results of Howard-Jones and Rosen (1993), who used only diotic stimulation and found evidence of asynchronous glimpsing for two bands, but not for greater numbers of bands.

Recall that in the Async-Δ-EVEN and Async-Δ-ODD conditions, the non-speech ear receives complementary bands of noise that are modulated out-of-phase relative to the masker bands in the signal ear. The monotic control conditions (Async-L-ODD and Async-R-EVEN) allow us to assess whether these noise bands, contralateral to the speech bands, had a masking effect. A masking effect did occur, with the monotic control conditions producing better SNRs than their respective dichotic controls at eight bands (see Table I); for example, performance in the Async-R-EVEN

condition is 5.8 dB better than the condition in which the contralateral masker is present (8-band, Async-Δ-EVEN). This difference may be related to the presence of noise in the opposite ear creating cross-ear interference, a possibility that will be addressed in the discussion.

C. Discussion

1. Energetic masking release

Howard-Jones and Rosen (1993) showed that asynchronous glimpsing of speech in asynchronous AM maskers is possible for small numbers of bands. The current study showed that presenting odd-numbered bands to one ear and even-numbered bands to the other ear improved the ability of the listener to identify the target speech. According to our hypothesis, this was a direct result of the elimination of peripheral masking spread that arose from neighboring bands in the monotic presentation. Specifically, the monotic masker bands in the "on"-phase likely introduced energetic masking into the neighboring spectral regions that were associated with the "off"-phase of modulation. This would have been especially likely for frequency regions above the upper cutoff of a masking band (Wegel and Lane, 1924). By presenting the alternating bands to opposite ears, the current study eliminated the effect of masking spread, and the result was an average of 7.3 dB more release from masking in the Async-Δ conditions compared to Async-R conditions. Since performance in the Async-Δ conditions was better than that associated with the dichotic control conditions, it is argued that listeners were integrating speech information across frequency and across ears, taking advantage of regions of high SNR distributed across frequency. This constitutes asynchronous glimpsing.

An alternate interpretation of the better thresholds in Async-Δ conditions than dichotic controls is that listeners are simply selecting the better subset of bands (even or odd) to attend to on a trial-by-trial basis in the Async-Δ conditions. This might be a good strategy if the critical information necessary to identify some consonants were better represented in even bands and the information necessary to identify other consonants were better represented in the odd bands. If listeners used different subsets of bands on a trial-by-trial basis, then performance would suffer in the control conditions due to elimination of one set of bands. While it is theoretically possible that listeners made use of information in different subsets of bands in the Async-Δ condition, there are two considerations that make this unlikely. First, previous data for the stimulus recordings used here suggest that the information necessary for correct identification is relatively uniformly distributed across odd- and even-numbered bands for individual consonants (Buss *et al.*, 2004). Second, consonant confusion matrices were analyzed for all conditions in the present study, and it was determined that while individual consonants were identified with varying accuracy, there was no evidence of consistently different error patterns in the just odd and just even control conditions. Interestingly, there was no evidence of a difference in error patterns between the Async-Δ and dichotic control conditions. These considerations strongly favor the interpretation that listeners were integrating across time and frequency in the Async-Δ conditions.[2]

2. Contralateral masking

It should be noted that as the number of bands increased, and consequently the bandwidth of each band narrowed, performance in the Async-Δ conditions decreased relative to the Sync conditions. This begs the question of what constraints other than spread of masking are placed on the listener when the masker was asynchronously modulated. An important point to consider is the effect that contralateral masking may have had in the Async-Δ conditions and their controls. Specifically, contralateral maskers may have introduced masking at a higher perceptual level. This effect could be related to findings in the literature described as central masking (Martin *et al.*, 1965; Martin and Digiovanni, 1979) or informational masking (Brungart and Simpson, 2002). Frequency effects have been observed in central masking, such that contralateral maskers are more effective when they are spectrally close to the target frequency (Zwislocki *et al.*, 1968). It is possible that central masking was greater for larger numbers of narrower maskers due to spectral proximity.

In a study by Brungart and Simpson (2002), listeners were found to have greater difficulty identifying monotic speech when it was masked by a dichotic speech competitor than when the competing speech was only in the ipsilateral ear. This effect disappeared when the contralateral ear (i.e., the opposite ear from the target speech) was presented with steady-state noise, a result interpreted as indicating that the contralateral competition requires a stimulus qualitatively similar to the target to cause a disruption in speech segregation. While the present study did not use competing speech as maskers, the maskers were spectro-temporally more complex than steady-state or even synchronous AM noise. The data show that identifying speech with only half the bands presented to a single ear was less difficult in the monotic controls than in dichotic controls (SRTs in the 8-band, Async-L-ODD and Async-R-EVEN conditions are better than the 8-band, Async-Δ-ODD and Async-Δ-EVEN conditions by 2.9 and 5.8 dB, respectively); this is consistent with an interpretation that the addition of the contralateral, opposite-phase masker in the dichotic controls greatly reduced unmasking due to central effects.

One possible way to conceive of across-ear interference is in terms of perceptual "miscuing" related to the phase of masker modulation. Buus (1985) proposed that the temporal envelope of a masker could alter perceptual weights in signal detection, such that more weight was applied during noise modulation minima (where the SNR was relatively good) and less weight during modulation maxima. It is possible that a related form of perceptual weighting contributes to the results of the present experiment. In the dichotic control conditions, when the masker was at a minimum in the signal ear, the masker was at a maximum in the contralateral ear. It is possible that the presence of masker peaks in the contralateral ear acted to reduce the weight given to the epochs of masker minima in the signal ear. This effect is not necessarily dependent on the presentation type (e.g., monotic or

dichotic), so it follows that performance in the monotic asynchronous AM conditions in the present study and in diotic asynchronous AM conditions in Howard-Jones and Rosen's study may have also been detrimentally affected by miscuing.

3. Spectro-temporal integration in asynchronous monotic AM

The data pattern in Fig. 2 could give the impression that there is no asynchronous glimpsing for two bands in the Async-R condition. Such a result would be inconsistent with the findings of Howard-Jones and Rosen (1993). One reason for this apparent discrepancy could be the nature of the control conditions used in these two studies. Whereas the dichotic control conditions in the present study did not have a signal in complementary spectral regions at a single ear, the control conditions of Howard-Jones and Rosen presented steady maskers in the complementary spectral regions. Inclusion of steady maskers in the previous study could have elevated thresholds via the introduction of spread of masking. Although elimination of masking spread in the control conditions was desirable for estimation of asynchronous glimpsing in the Async-Δ conditions of the present study, it may not provide an appropriate reference for asynchronous glimpsing in the Async-R conditions.

Supplemental data were collected to determine whether the presence of steady maskers in the control condition is an important procedural factor. Data were collected on four listeners (two original participants and two practiced, new participants). Monotic presentations included an unmodulated noise condition and two control conditions incorporating modulated and steady masker bands, following the procedures of Howard-Jones and Rosen (1993). Control conditions were based on either two or four bands, and they included either even or odd numbered stimulus bands (speech and AM noise), as well as bands of steady noise in the spectral regions of the complementary odd or even bands (see Fig. 1; Async-L-ODD* and Async-R-EVEN*). The mean SRT in the unmodulated condition was lower in the supplementary data than in the primary experiment (2.8 dB), an effect we attribute to individual differences. Despite good performance in the baseline condition for these listeners, the SRT in the better control condition was worse than that in the associated Async-Δ better controls of the main experiment, with differences of 7.6 dB (2-band) and 1.6 dB (4-band). Performance in the Async-R conditions was better than that in the steady-band control conditions, an effect of 6.6 dB (2-band) and 6.9 dB (4-band); these results indicate substantial glimpsing in the Async-R condition when spread of masking is incorporated into the control condition.

Although some apparent differences between the outcomes of the present experiment and the experiment of Howard-Jones and Rosen appear to be accounted for by the different control conditions, a difference in results for the Async-R condition is harder to explain. Whereas we found that the Async-R SRTs for 8 and 16 bands were better than for unmodulated noise, Howard-Jones and Rosen found that the SRTs in asynchronously modulated noise were no better for 8 and 16 bands than for unmodulated noise. One factor that could account for this difference is presentation level. The stimulation level is not reported in Howard-Jones and Rosen (1993), but if a higher level were used than in the present experiment, this would result in greater spread of masking and less ability to benefit from modulation with large numbers of bands.

III. EXPERIMENT 2

Results from experiment 1 showed evidence of asynchronous integration in the dichotic stimulus conditions. Additionally, release from masking relative to the unmodulated control condition was as much as 22.3 dB in the dichotic asynchronous condition, just slightly below the roughly 23-dB release for the two Sync conditions. Because there was an additional benefit of having both sets of masked speech bands in the Async-Δ conditions over the dichotic controls—with between 4.1 and 7.2 dB greater masking release, depending on the number of frequency bands—it seems unlikely that listeners used just the bands presented to a single ear.

The second experiment tested the robustness of the dichotic benefit when more detailed speech information is required in order to make a correct response. The response set-size for speech identification can change the benefit of masker AM due to differences in the amount of detail needed to perform the task. In a study by Buss and colleagues (2009), masking release for words in synchronous AM noise was found to differ depending on the set-size of the speech recognition task. When listeners were asked to identify a target word without constraints, masking release was smaller than when they were asked to select from among three alternatives: In one set of conditions, masking ranged from 8.7 dB (open-set) to 14.5 dB (closed-set) for synchronous 10-Hz amplitude modulation. The authors argued that reducing constraints on the response alternatives increases the amount of information necessary to perform well on the task, and therefore reduces the ability to do well based on sparse glimpses of the speech. It follows that if the set-size is manipulated for the identification tasks, listeners will have greater difficulty in the conditions with the least acoustic speech information.

Experiment 2 examined asynchronous glimpsing as a function of response set-size. Due to the importance of speech redundancy in an open-set task and the paucity of information present in each subset of bands (just odd and just even), a greater reliance on integration across time and frequency in the asynchronous AM condition is expected. As in experiment 1, it was expected that the elimination of masking spread would produce a general benefit for dichotic stimulation compared to monotic stimulation, with more evidence of glimpsing in an open-set task than a closed-set word recognition task.

A. Method

1. Listeners

Ten listeners participated in experiment 2, and all met inclusion criteria stated in experiment 1. Five listeners were tested in the open-set protocol and five in the closed-set

protocol. Four of the ten listeners had been previously tested on experiment 1.

2. Stimuli

The speech material for experiment 2 was a set of 500 consonant-nucleus-consonant (CNC) words (Peterson and Lehiste, 1962), spoken by an adult male with an American accent. Recordings were 444–992 ms, with a mean duration of 744 ms. The sampling rate was 24414 Hz, and all signals were passed through an 8000-Hz second order Butterworth low-pass filter. Recordings were digitally scaled to equal-rms level across tokens. Speech tokens were up-sampled to 48828 Hz to conform to hardware specifications. Filtering the speech into 2, 4, 8, or 16 frequency bands was performed using the same methods described above for experiment 1.

All masking stimuli were identical to those in experiment 1 with the exception that stimulus duration was equal to the longest possible speech token plus 300 ms (1292 ms total duration). All stimuli could be presented monotically (L or R) or dichotically (Δ). Dichotic stimulation presented the odd-numbered bands to the left ear and the even-numbered bands to the right ear.

3. Procedure

The level of the speech target was fixed, and the masker level was varied using an adaptive "2-up-1-down" procedure to determine the SRT associated with 71% correct (Levitt, 1971). The same hardware, target sound level, masker level step size, and listening environment were used as in the first experiment. Each SRT estimate was computed as the mean masker level at the last of 10 of 12 track reversals, and test conditions were randomly arranged to avoid order effects.

For this experiment, two protocols were employed. The first protocol was a closed-set, 4-alternative-forced choice identification task. The listener responded by clicking a button corresponding to the presented CNC word from a display of four choices, including the target and three randomly selected foils. The second protocol was an open-set, free response identification task. The listener responded by repeating the target word aloud; at that point the listener was visually presented with the correct response and prompted to score his or her response as correct or incorrect using buttons displayed on the computer screen. An experimenter monitored the experimental session, including spot checks for correct self-scoring. As in experiment 1, there were 21 experimental conditions: one reference condition (Unmod), two synchronous AM conditions (Sync-R and Sync-Δ with eight bands), and two asynchronous conditions (Async-R and Async-Δ) with 2, 4, 8, or 16 bands. Dichotic controls were tested for each Async-Δ condition, and there were additional 8-band monotic controls (Async-L-ODD and Async-R-EVEN), as in experiment 1 (see Fig. 1 for reference). A minimum of three threshold estimates was obtained in all conditions. In the event that thresholds in a particular condition varied by more than 3 dB, an additional threshold was collected. This occurred for 7–15 of the 21 conditions in the closed-set task and for 14–20 of the 21 conditions in the open-set task, depending on the listener. Overall testing time was roughly 4 h per protocol, spread out over three separate 1–1.5 h sessions.

B. Results and discussion

1. Closed-set speech reception thresholds

Figure 3 (left panel) shows the mean SRTs (in dB SNR) for each masker condition in the closed-set protocol relative to the reference condition and follows the same plotting convention as in Fig. 2 (see also Table II). The reference (Unmod) condition had a mean SRT of –9.3 dB SNR. Note that this threshold is better than that obtained in the unmodulated noise condition of experiment 1 (−1.9 dB SNR), consistent with an interpretation that the present four-alternative word task was easier than the 12-choice consonant task of experiment 1, despite the fact that this closed-set task tracked a higher percent correct (71% vs 50%). Release from masking in the closed-set tasks was not significantly different for the Sync-R and Sync-Δ conditions ($t_4 = 1.03$, $p = 0.36$), with a mean of 22.7 dB. This value is similar to that observed in experiment 1. The Async-Δ masking release ranged from 23.2 to 18.4 dB as band number increased, and Async-R masking release ranged from 17.7 to 5.1 dB as band number increased. Once again, as Howard-Jones and Rosen (1993) observed, an increase in band number in the Async-R conditions reduced the overall performance relative to the synchronous conditions, though, in the present study, a masking release in the Async-R conditions was obtained for all numbers of bands. Submitting the Async-R and Async-Δ thresholds to a two-way ANOVA with two levels of condition and four levels of number of bands confirmed a main effect of condition ($F_{1,4} = 27.1$, $p < 0.01$), a main effect of number of bands ($F_{3,12} = 20.1$, $p < 0.001$), and a significant interaction ($F_{3,12} = 6.57$, $p < 0.01$). The interaction was due to the relatively steep increase in SRT for the Async-R

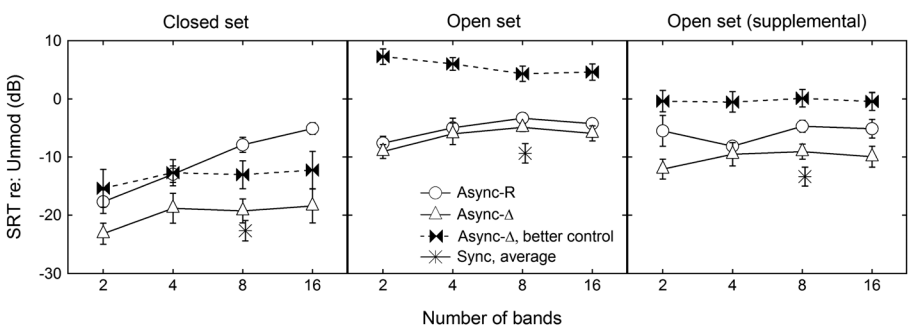

FIG. 3. Mean SRTs for the closed-set (left panel) and open-set (middle panel) protocols in experiment 2 are plotted for modulated noise conditions relative to the unmodulated condition. Supplementary data for an open-set protocol with target presented at 55 dB SPL are plotted in the right panel. Plotting style follows from Fig. 2. Error bars indicate standard error of the mean ($n = 5$).

TABLE II. Mean SRTs (in dB SNR) from experiment 2 are reported for each stimulus condition. The standard error of the mean ($n = 5$) is shown in parentheses below the associated mean. For the dichotic control conditions, the condition associated with better performance is indicated by an asterisk for each number of bands. Recall that control conditions included only half of the speech bands of the associated Async condition; for example, the 8-band control included only four bands of speech.

Conditions		Unfiltered	Number of bands			
			2	4	8	16
Closed-set	Unmod	−9.2 (0.2)				
	Sync-R	−32.2 (1.6)				
	Sync-Δ				−31.7 (1.8)	
	Async-R		−27.0 (1.9)	−22.2 (1.5)	−17.2 (1.4)	−14.4 (1.1)
	Async-Δ		−32.5 (1.7)	−28.1 (2.5)	−28.6 (2.0)	−27.7 (2.8)
	Async-Δ-ODD		−24.7* (3.3)	−21.2 (2.7)	−22.3* (2.4)	−21.5* (3.2)
	Async-Δ-EVEN		−20.8 (2.8)	−22.0* (2.2)	−20.5 (4.0)	−20.4 (1.6)
	Async-L-ODD				−26.9 (3.4)	
	Async-R-EVEN				−25.6 (2.5)	
Open-set	Unmod	5.7 (1.3)				
	Sync-R	−3.6 (3.0)				
	Sync-Δ				−3.7 (2.7)	
	Async-R		−1.9 (1.2)	0.7 (1.5)	2.4 (1.1)	1.5 (1.7)
	Async-Δ		−3.3 (2.2)	−0.3 (1.5)	1.0 (1.4)	−0.2 (1.7)
	Async-Δ-ODD		24.4 (0.3)	17.2 (1.6)	11.2 (2.9)	10.8 (2.4)
	Async-Δ-EVEN		13.1* (1.6)	11.7* (2.1)	10.3* (2.4)	10.3* (2.4)
	Async-L-ODD				9.0 (2.7)	
	Async-R-EVEN				10.4 (3.5)	
Supplemental data	Unmod	6.8 (0.9)				
	Sync-R	−7.6 (1.8)				
	Sync-Δ				−5.6 (1.6)	
	Async-R		1.3 (2.4)	−1.3 (0.7)	2.1 (1.82)	1.7 (0.9)
	Async-Δ		−5.3 (2.2)	−2.7 (1.6)	−2.3 (0.7)	−3.2 (1.4)
	Async-Δ-ODD		32.8 (1.2)	10.0 (1.2)	8.5 (1.8)	6.4* (1.6)
	Async-Δ-EVEN		6.4* (1.9)	7.3* (1.5)	6.9* (1.3)	7.9 (1.0)
	Async-L-ODD				7.5 (2.2)	
	Async-R-EVEN				3.5 (1.6)	

conditions as band number increased beyond four, compared to the relatively flatter function of thresholds for the Async-Δ conditions. Just as the data from experiment 1 suggested, there was a clear increase in masking release for dichotic asynchronous AM conditions (Async-Δ) compared to the corresponding monotic conditions (Async-R), with a mean difference of 5.5 to 13.3 dB.

As in experiment 1, the present Async-Δ data showed a clear benefit for dichotic asynchronous masker presentation over the dichotic controls (see Fig. 3). Submitting the Async-Δ and better control thresholds to a two-way ANOVA with two levels of condition and four levels of number of bands confirmed a main effect of condition ($F_{1,4} = 73.4$, $p = 0.001$), a main effect of number of bands ($F_{3,12} = 9.83$, $p = 0.001$), and no significant interaction ($F_{3,12} = 0.37$, $p = 0.77$). Averaged over the band number conditions, SRTs for the Async-Δ condition were 6.6 dB better than for the better dichotic control. By providing the listener with more speech information in the Async-Δ conditions, performance was better than when only the odd or even speech bands were present. This provides evidence for integration across ears and frequency bands. For the 8-band conditions, thresholds were 4.5 and 5.1 dB better in the monotic than the dichotic control conditions (odd and even, respectively). These results indicate that including contralateral masker bands with out-of-phase modulation hurts performance, as in experiment 1.

Whereas results from the Async-Δ conditions likely reflect asynchronous glimpsing, performance in the Async-R conditions was comparable to or worse than that in the better Async-Δ control condition. Better performance in the control conditions likely reflects the benefits of eliminating spread of masking. Recall that the dichotic control conditions presented either the odd- or even-numbered speech and AM noise bands to one ear, and the remaining masker bands to the other ear. This dichotic masker presentation would improve the peripheral representation of speech bands during masker minima. For the closed-set tasks, this information was sufficient to support performance that was comparable to or better than that in the Async-R conditions, when all speech bands were present.

2. Open-set speech reception thresholds

Overall, thresholds were poorer in the open-set than the closed-set conditions. The mean SRT in the reference (Unmod) condition was 5.7 dB SNR, consistent with the relative difficulty of the task. Results in the two Sync conditions were not significantly different ($t_4 = 0.06$, $p = 0.95$),

with an average of 9.4 dB masking release (Fig. 3, middle panel). Table II also shows the mean SRTs (in dB SNR). More masking release with synchronous masker modulation for a closed-set task than an open-set task has precedent in the literature (Buss *et al.*, 2009) and may be related to the finding of greater masking release for conditions associated with better performance in the baseline condition (Bernstein and Brungart, 2011). In light of the reduced masking release in the synchronous AM conditions, it is not surprising that masking release in the Async conditions was also markedly reduced when compared to the closed-set protocol. For the Async-R conditions, masking release ranged from 7.6 to 3.3 dB across the different band number conditions. In comparison, for the Async-Δ conditions, masking release ranged from 9.0 to 4.7 dB. A two-way ANOVA with two levels of condition and four levels of number of bands yielded a main effect of number of bands ($F_{3,12} = 4.84$, $p < 0.05$), no main effect of condition ($F_{1,4} = 1.20$, $p = 0.33$), and no interaction ($F_{3,12} = 0.03$, $p = 0.99$).

Inspection of Fig. 3 shows a benefit for dichotic asynchronous masker presentation over the dichotic controls. Submitting the Async-Δ and better dichotic control thresholds to a two-way ANOVA with two levels of condition and four levels of number of bands confirmed a main effect of condition ($F_{1,4} = 203.2$, $p < 0.001$), no effect of number of bands ($F_{3,12} = 0.22$, $p = 0.88$), and a significant interaction ($F_{3,12} = 4.28$, $p < 0.05$). That interaction was due to a greater difference between Async-Δ and control conditions for the lower band numbers than the higher band numbers. Mean SRTs in the Async-Δ conditions were on average 12.1 dB less than the better dichotic controls. By providing the listener with more speech information in the Async-Δ conditions, performance was better than when only the odd or even speech bands were present. This was evidence for integration across ears and frequency bands. Thresholds were similar in the monotic and dichotic 8-band control conditions, which differed by 2.2 dB or less. These results are consistent with the notion that a relatively difficult speech task such as open-set word recognition requires a great deal of speech detail and redundancy (Buss *et al.*, 2009), which these control conditions lacked.

In contrast to the closed-set data, the open-set task performance in the Async-R conditions was consistently superior to that in the dichotic controls. This is likely due to the fact that whereas just the even or just the odd bands supported relatively good performance in the closed-set task, this was not the case in the open-set task. These results indicate that asynchronous glimpsing occurred for the Async-R despite the presence of spread of masking.

3. Asynchronous glimpsing

Whereas masking release is defined relative to the Unmod baseline, asynchronous glimpsing is the ability to combine information across frequency regions containing asynchronously modulated masker bands. The magnitude of asynchronous glimpsing was calculated as the difference between thresholds in the Async-Δ and the better of the two dichotic control conditions (Async-Δ-EVEN and Async-Δ-ODD). Mean values of asynchronous glimpsing are plotted as a function of the number of bands in Fig. 4, with symbol style reflecting response conditions as defined in the legend. In open-set data, glimpsing ranged from 16.4 to 9.4 dB depending on the number of bands (open symbol). Contrast those numbers to the case for the closed-set protocol, in which glimpsing ranged from only 7.8 to 6.1 dB (filled symbol). A repeated-measures ANOVA was performed to evaluate the effect of response protocol on the magnitude of asynchronous glimpsing. There were four within-subjects levels of band and a between-subject factor of protocol. Results confirmed a main effect of protocol ($F_{1,8} = 23.0$, $p = 0.001$), a main effect of the number of bands ($F_{3,24} = 3.69$, $p < 0.05$), but no interaction between number of bands and protocol ($F_{3,24} = 1.39$, $p = 0.27$). This reflects the fact that there is greater evidence of integration across time and frequency in the open set than the closed set protocol, but glimpsing was reduced similarly between protocols as the number of bands increased. Therefore, while performance in the modulated masker condition was worse overall in the open-set protocol, the magnitude of asynchronous glimpsing was significantly greater than in the closed-set protocol. This task effect is consistent with the hypothesis that asynchronous glimpsing is likely to be more pronounced when detailed speech cues are needed, as in the open-set task.

4. Supplemental data on possible level effects

While this experiment demonstrated a significant difference between monotic and dichotic asynchronous AM conditions in the closed-set task, no significant difference was observed in the open-set task. The explanation for this may be based in the nature of masking spread at high presentation levels, such that spread of masking increases particularly on the high side of the masker (Wegel and Lane, 1924; Moore *et al.*, 1998). In this study, presentation for the target speech was chosen so that the masker was loud but comfortable in the easiest condition. Since the Sync conditions were

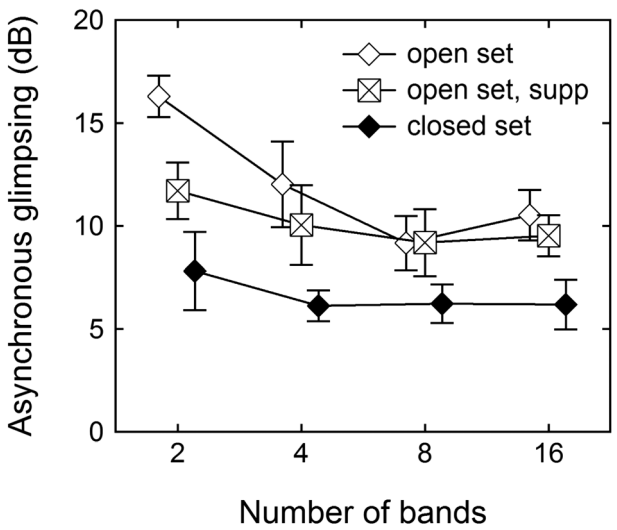

FIG. 4. Asynchronous glimpsing in experiment 2, calculated as the difference in SRT between the dichotic asynchronous condition and the better of two dichotic control conditions (Async-Δ-EVEN and Async-Δ-ODD). Symbols indicate the test protocol, as defined in the legend.

associated with the best performance (i.e., highest masker level), these conditions essentially dictated the target level, 45 dB SPL. Once this level was chosen in experiment 1, it was kept constant for experiment 2 in order to facilitate comparison of datasets. As a result, in the open-set protocol of experiment 2, which was more difficult than the other protocols, it is likely that masker levels did not reach high enough intensities to produce large effects of spread of masking. If the Async-R condition was not substantially affected by spread of masking, it is logical that separation of the stimulus bands between the ears (Async-Δ) would not have the beneficial outcome that it had in other conditions.

This interpretation was evaluated by collecting additional data in the open-set task of experiment 2, but with the target level increased from 45 to 55 dB SPL. Mean SRTs relative to baseline from five naïve subjects are plotted in Fig. 3 (right panel) and reported in Table II (in dB SNR). Increasing the target level by 10 dB increased the baseline threshold to 48.2 dB SPL, or 6.8 dB SNR. As shown in the figure, masking release was greater for dichotic than monotic conditions, consistent with an interpretation that an increase in overall level can result in greater spread of masking and, therefore, a dichotic listening advantage. A two-way ANOVA confirmed a main effect of condition between Async-R and Async-Δ ($F_{1,4} = 27.1$, $p < 0.01$), but no main effect of the number of bands ($F_{3,12} = 0.87$, $p = 0.48$), and no interaction between condition and number of bands ($F_{3,12} = 0.82$, $p = 0.51$). Compared to the asynchronous dichotic condition (Async-Δ), thresholds were on average 4.3 dB poorer in the Async-R condition. The 8-band Async-Δ threshold is on average 7.8 dB better than the monotic control conditions. Asynchronous glimpsing calculated for these supplemental data is shown in Fig. 4. While there tends to be reduced asynchronous glimpsing in the supplemental compared to the primary open-set data, glimpsing tended to be greater in the supplemental open-set data than in the closed-set data. This shows that while the open-set task relies more heavily on asynchronous glimpsing than the closed-set task, this task effect is reduced at higher presentation levels, in which speech cues may be more salient in the dichotic controls.

While the supplemental open-set data are interpreted as reflecting greater spread of masking, it is possible that increased audibility of low-level speech cues could have played a role in these results. Audibility of speech does not ensure that all speech cues are audible, and it is possible that some low-level speech cues were audible at 55 dB SPL but not 45 dB SPL. This possibility is supported by the finding that audibility of cues that are 28 dB below the peak rms level can affect performance (Studebaker *et al.*, 1999). While these low-level cues may not be critical for speech presented in quiet or in steady noise, they may become critical to recognition in modulated noise, where the cues available are temporally sparse. If audibility were an important factor in the performance associated with these stimuli, masking release in the synchronous conditions would be greater at the higher than the lower presentation level. This was the case: Averaging across Sync-R and Sync-Δ conditions, masking release was 9.4 and 13.4 dB for the 45 and 55 dB SPL speech levels, respectively. While low-level cues may have played a role in the better Sync performance, it is unclear how the increased audibility of low-level cues could have reduced the difference between the Async-R and Async-Δ conditions. It is a logical possibility that audibility could impact asynchronous glimpsing. However, it is more parsimonious to argue that low- and high-level speech cues are integrated across time and frequency in a similar way, and that spread of masking is the critical difference between the primary and supplemental open-set data.

IV. GENERAL DISCUSSION

The present study tested the idea that asynchronous glimpsing of speech could be aided by dichotic presentation, in which neighboring frequency bands are separated between the ears. It is important to understand how listeners integrate partial speech information across time and frequency since real-world acoustic environments are not always spectro-temporally uniform. Experiment 1 extended the work of Howard-Jones and Rosen (1993) by presenting asynchronously masked speech dichotically, whereas the previous work had only presented stimuli diotically. By presenting stimuli dichotically, with even and odd frequency bands separated to opposite ears, peripheral spread of masking was avoided. The results show that asynchronous glimpsing of speech is achievable for 2–16 bands, and poorer performance in dichotic controls than dichotic asynchronous conditions provided evidence against the possibility that listeners were relying on information in just the even or just the odd numbered bands.

While the dichotic presentation of the asynchronous AM masker improved asynchronous glimpsing, the data of experiment 1 show a significant decline in performance as band number increased regardless of whether stimuli were presented monotically or dichotically. This result is similar to the pattern of results seen by Howard-Jones and Rosen. Assuming that the effects of masking spread have been eliminated in the Async-Δ condition, it is unclear why performance would decline at higher band numbers. This is especially interesting given the results of Buss *et al.* (2004), showing that performances remained relatively consistent across all band numbers when the speech was modulated out-of-phase. Further, the difference between Async-Δ and Async-R thresholds should increase with increasing number of bands, to the extent that spread of masking has a larger effect on the Async-R performance with larger numbers of bands. This trend was significant only for the closed-set data of experiment 2. One possibility is that listeners had greater difficulty in the asynchronous condition because masker minima in the even bands coincided with masker maxima in the odd bands, and vice versa. The masker peaks may have reduced perceptual weights associated with speech information in the coincident masker dips, thereby limiting benefit related to asynchronous glimpsing (Buus, 1985). Such reduced weights would limit the ability to benefit from reduced spread of masking. To be fully consistent with the data, however, such an effect would have to depend on the speech material (VCVs and CNC words) and/or the listener's task; it is unclear why that might be the case. One reason there may be differences in effects with different speech

material lies in the relative temporal width of important speech cues between VCVs and CNC words. Specifically, the VCVs require a much shorter temporal glimpse for consonant recognition, whereas the CNC words rely more heavily on the longer vowel information. This difference in required temporal glimpse for accurate speech recognition may interact with the type of presentation (monotic or dichotic), but further experiments would be necessary to understand why.

Experiment 2 was designed to assess the nature of asynchronous AM masking as a function of speech identification task. Previous research has indicated that the amount of unmasking is influenced by the complexity of speech information required to perform a speech recognition task. Particularly, Buss *et al.* (2009) recently observed that manipulating the set-size of speech identification tasks can greatly alter the amount of the target signal information required to perform well. In a speech recognition task using CNC words, Buss *et al.* found masking release for synchronous AM maskers in a closed-set task was roughly 7 dB greater than in an open-set task. This is broadly consistent with our results in experiment 2, which showed unmasking to be 13.4 dB greater in the closed-set task than the open-set task for synchronous AM maskers when the target speech is played at 45 dB SPL. For asynchronous AM masking of a 45-dB-SPL signal, our experiment showed that a difference between monotic and dichotic listening was present only in the closed-set protocol. The interpretation of this difference is confounded by the fact that masker level was lower at threshold in the open-set protocol due to greater difficulty of the task. Supplementary data, in which the target level was increased by 10 dB, did show a benefit for dichotic listening. This is consistent with the idea that spread of masking hinders monotic asynchronous glimpsing at relative high stimulus level, where spread of masking is largest.

The results of experiment 2 and the supplementary data provide insight into the effect of speech task set size on asynchronous glimpsing. While overall unmasking declined in the open-set task, there was greater evidence of asynchronous glimpsing in the open-set than the closed-set task (Fig. 4). An increase in asynchronous glimpsing was hypothesized for the open-set protocol due to greater requirements on the amount of speech information necessary to perform the task. Since asynchronous glimpsing was calculated as the difference between thresholds in the dichotic asynchronous AM conditions and those of the better dichotic control, insufficient cues in either the even or odd bands alone would have impacted the dichotic controls more severely in the open-set than the closed-set task. In fact, the results showed that the smallest estimate of asynchronous glimpsing (in dB) in the open-set task was greater than the largest estimate of glimpsing in the closed-set task. This outcome supported our hypothesis that the open-set condition would be associated with reduced masking release and increased evidence of asynchronous glimpsing.

V. CONCLUSIONS AND POSSIBLE CLINICAL RELEVANCE

The present study tested whether asynchronous glimpsing in Howard-Jones and Rosen's (1993) study was limited by the peripheral spread of masking, particularly for large numbers of narrow bands. By presenting even bands and odd bands of speech and asynchronous AM maskers to opposite ears, we have shown that significantly greater release from masking is possible with dichotic presentation. A benefit of asynchronous masker modulation is obtained even when maskers are filtered into as many as 16 bands. However, it should be noted that performance declined as the number of bands increased for both monotic and dichotic asynchronous conditions in experiment 1. The present data do not allow an unambiguous account of this effect, but it is possible that it may be related to some detrimental effect of miscuing. Additionally, while no benefit of dichotic over monotic presentation of asynchronous stimuli was observed in the primary open-set task of experiment 2, it is likely that the low masker level at threshold played a role in this result, as supported by the supplementary data.

This study provides new evidence that normal-hearing listeners are able to integrate speech information asynchronously across time and frequency. The current maskers are predictable in their spectro-temporal structure, and therefore do not reflect the randomness of many natural masking environments. Nevertheless, this study has possible implications for hearing aid design for those with hearing impairment. For example, bilateral auditory prostheses could implement processing strategies to ameliorate the disruptive effects of masking spread between neighboring frequency regions by splitting even and odd numbered bands to opposite ears (Franklin, 1981; Lunner *et al.*, 1993). Since effects of masking spread might be even more pronounced in hearing-impaired listeners due to reduced frequency selectivity (Florentine *et al.*, 1980), further study would be required to evaluate the effect of band number on asynchronous glimpsing of dichotic information in hearing-impaired listeners.

ACKNOWLEDGMENTS

This work was supported by NIH NIDCD R01 DC000418. A subset of these data was presented at the 2011 mid-winter meeting of the Association for Research in Otolaryngology. We thank Peter Gordon and Neil Mulligan for their helpful comments. We would also like to thank Laurent Demany and two anonymous reviewers for their useful remarks on this manuscript.

[1]In contrast to the study by Howard-Jones and Rosen (1993), where stimulation was diotic, the present study used monotic presentations. This was intended to be a better control for the dichotic conditions based on the fact that a diotic stimulus inherently incorporates two equal presentations, whereas the dichotic stimulus presents only a single representation of each signal band. While the equivalent diotic presentations would have been useful to compare the Howard-Jones and Rosen results to the results obtained here, the primary focus of the current study was to test dichotic stimulation, and to determine whether dichotic performance exceeded that supported by just a subset of bands (i.e., even only or odd only).

[2]An unforeseen percept was noted by some listeners for the Async-Δ conditions; specifically, it was possible to perceive the masker as spatially separated from the target speech, which may be related to the spatial percept obtained in contralateral induction studies (Warren and Bashford, 1976). It is not clear whether this percept might have contributed to masking release. In order to get some idea of the magnitude of a spatial masking release effect for these stimuli, the three modulation types (Unmod,

Sync, and Async) were tested in three typical masking-level difference (Hirsh, 1948) configurations: NmSm (speech and noise presented monotically), NoSm (noise presented diotically and speech presented monotically), and NuSm (independent noises presented to the two ears and speech presented monotically). In the latter two conditions, the speech and masker were associated with different spatial percepts, but performance was similar to that in the NmSm case. There was no difference between the NuSm and NmSm thresholds, and the NoSm threshold was only 3.5 dB better. Since the data of experiment 1 show 5.1–8.5 dB more release from masking in the Async-Δ conditions compared to the Async-R conditions, it is unlikely that such unmasking is solely due to the perceived spatial separation.

ANSI (**1994**). *ANSI S1.1-1994, American National Standard Acoustical Terminology* (American National Standards Institute, New York).

Bernstein, J. G., and Brungart, D. S. (**2011**). "Effects of spectral smearing and temporal fine-structure distortion on the fluctuating-masker benefit for speech at a fixed signal-to-noise ratio," J. Acoust. Soc. Am. **130**, 473–488.

Brungart, D. S., and Simpson, B. D. (**2002**). "Within-ear and across-ear interference in a cocktail-party listening task," J. Acoust. Soc. Am. **112**, 2985–2995.

Buss, E., Hall, J. W., III, and Grose, J. H. (**2003**). "Effect of amplitude modulation coherence for masked speech signals filtered into narrow bands," J. Acoust. Soc. Am. **113**, 462–467.

Buss, E., Hall, J. W., III, and Grose, J. H. (**2004**). "Spectral integration of synchronous and asynchronous cues to consonant identification," J. Acoust. Soc. Am. **115**, 2278–2285.

Buss, E., Whittle, L. N., Grose, J. H., and Hall, J. W., III (**2009**). "Masking release for words in amplitude-modulated noise as a function of modulation rate and task," J. Acoust. Soc. Am. **126**, 269–280.

Buus, S. (**1985**). "Release from masking caused by envelope fluctuations," J. Acoust. Soc. Am. **78**, 1958–1965.

Dirks, D. D., and Bower, D. (**1970**). "Effect of forward and backward masking on speech intelligibility," J. Acoust. Soc. Am. **47**, 1003–1008.

Florentine, M., Buus, S., Scharf, B., and Zwicker, E. (**1980**). "Frequency selectivity in normally-hearing and hearing-impaired observers," J. Speech Hear. Res. **23**, 646–669.

Franklin, B. (**1981**). "Split-band amplification: A HI/LO hearing aid fitting," Ear Hear. **2**, 230–233.

George, E. L., Festen, J. M., and Houtgast, T. (**2006**). "Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **120**, 2295–2311.

Gnansia, D., Jourdes, V., and Lorenzi, C. (**2008**). "Effect of masker modulation depth on speech masking release," Hear. Res. **239**, 60–68.

Grose, J. H., and Hall, J. W., III (**1992**). "Comodulation masking release for speech stimuli," J. Acoust. Soc. Am. **91**, 1042–1050.

Hall, J. W., Grose, J. H., and Dev, M. B. (**1997**). "Signal detection and pitch ranking in conditions of masking release," J. Acoust. Soc. Am. **102**, 1746–1754.

Hall, J. W., Haggard, M. P., and Fernandes, M. A. (**1984**). "Detection in noise by spectro-temporal pattern analysis," J. Acoust. Soc. Am. **76**, 50–56.

Hirsh, I. J. (**1948**). "Binaural summation; A century of investigation," Psychol. Bull. **45**, 193–206.

Howard-Jones, P. A., and Rosen, S. (**1993**). "Uncomodulated glimpsing in 'checkerboard' noise," J. Acoust. Soc. Am. **93**, 2915–2922.

Kwon, B. J. (**2002**). "Comodulation masking release in consonant recognition," J. Acoust. Soc. Am. **112**, 634–641.

Li, N., and Loizou, P. C. (**2007**). "Factors influencing glimpsing of speech in noise," J. Acoust. Soc. Am. **122**, 1165–1172.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Lunner, T., Arlinger, S., and Hellgren, J. (**1993**). "8-channel digital filter bank for hearing aid use: Preliminary results in monaural, diotic and dichotic modes," Scand. Audiol. Suppl. **38**, 75–81.

Martin, F. N., Bailey, H. A. T., and Pappas, J. J. (**1965**). "The effect of central masking on threshold for speech," J. Aud. Res. **5**, 293–296.

Martin, F. N., and Digiovanni, D. (**1979**). "Central masking effects on spondee threshold as a function of masker sensation level and masker sound pressure level," J. Am. Aud. Soc. **4**, 141–146.

Miller, G. A., and Licklider, J. C. R. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**, 167–163.

Moore, B. C., Alcantara, J. I., and Dau, T. (**1998**). "Masking patterns for sinusoidal and narrow-band noise maskers," J. Acoust. Soc. Am. **104**, 1023–1038.

Nelken, I., Rotman, Y., and Bar Yosef, O. (**1999**). "Responses of auditory-cortex neurons to structural features of natural sounds," Nature **397**, 154–157.

Peters, R. W., Moore, B. C., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," J. Acoust. Soc. Am. **103**, 577–587.

Peterson, G. E., and Lehiste, I. (**1962**). "Revised CNC lists for auditory tests," J. Speech Hear. Disord. **27**, 62–70.

Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gwaltney, C. A. (**1999**). "Monosyllabic word recognition at higher-than-normal speech and noise levels," J. Acoust. Soc. Am. **105**, 2431–2444.

Summers, V., and Molis, M. R. (**2004**). "Speech recognition in fluctuating and continuous maskers: Effects of hearing loss and presentation level," J. Speech Lang. Hear. Res. **47**, 245–256.

Warren, R. M., and Bashford, J. A. (**1976**). "Auditory contralateral induction: An early stage in binaural processing," Percept. Psychophys. **20**, 380–386.

Wegel, R. L., and Lane, C. E. (**1924**). "The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear," Phys. Rev. **23**, 266.

Zwislocki, J., Buining, E., and Glantz, J. (**1968**). "Frequency distribution of central masking," J. Acoust. Soc. Am. **43**, 1267–1271.