Speech intelligibility among modulated and spatially distributed noise sources

John F. Culling[a] and Elizabeth R. Mansell
School of Psychology, Cardiff University, Tower Building, Park Place, Cardiff CF10 3AT, United Kingdom

(Received 28 May 2012; revised 1 February 2013; accepted 11 February 2013)

At a cocktail party, listeners are faced with multiple, spatially distributed interfering voices. The dominant interfering voice may change from moment to moment and, consequently, change in spatial location. The ability of the binaural system to deal with such a dynamic scene has not been systematically analyzed. Spatial release from masking (SRM) was measured in simple spatial scenes, simulated over headphones with a frontal speech source. For a single noise at 105°, SRM was reduced if that noise modulated (10 Hz square wave, 50% duty cycle, 20 dB modulation depth), but, for two noises in symmetrical locations, SRM increased if the noises were modulated in alternation, suggesting that the binaural system can "switch" between exploiting different spatial configurations. Experiment 2 assessed the contributions of interaural time and level differences as a function of modulation rate (1–20 Hz). Scenes were created using the original head-related impulse responses and ones that had been manipulated to isolate each cue. SRM decreased steeply with modulation rate. The combined effects of interaural time and level differences were consistent with additive contributions. The results indicate that binaural sluggishness limits the contribution of binaural switching to speech understanding at a cocktail party.
 [http://dx.doi.org/10.1121/1.4794384]

PACS number(s): 43.66.Dc, 43.66.Pn, 43.66.Ba [MAA] Pages: 2254–2261

I. INTRODUCTION

The benefit of spatial separation to speech intelligibility in continuous noise is now well understood. However, speech is often heard against multiple dynamic interferers, such as a babble of competing voices. These voices fluctuate in their dominance of the overall interfering complex, raising the possibility that the auditory system might address the cocktail-party problem by suppressing different interfering sounds at different points in time (Peissig and Kollmeier, 1997). The present experiment examined the potential role for such a mechanism using noise maskers in different virtual locations that were modulated in alternation.

Spatial release from masking (SRM) is the improvement in speech reception threshold (SRT) when a spatial separation is introduced. SRM in unmodulated continuous noise can be accurately predicted as a combination of better-ear listening and binaural unmasking (Zurek, 1993; Beutelmann and Brand, 2006; Lavandier and Culling, 2010; Beutelmann *et al.*, 2010; Wan *et al.*, 2010; Jelfs *et al.*, 2011). Better-ear listening involves taking advantage of the ear with the better signal-to-noise ratio (SNR), while binaural unmasking involves the combination of the stimuli at the two ears to suppress noise with a given interaural time delay (ITD), possibly via a cancellation mechanism (Durlach, 1963, 1972; Culling, 2007).

Intelligibility in modulated noise is less well understood. Intelligibility is usually improved by masker modulation (Miller and Licklider, 1950; de Laat and Plomp, 1983; Festen, 1987), and this benefit is often referred to as "masking release" or "dip listening." The effect can be predicted in limited circumstances by calculating the average Speech Intelligibility Index (ANSI, 1997) over a series of time windows (Rhebergen and Versfeld, 2005).[1] However, Stone *et al.* (2012) has argued that masking release occurs not because listeners selectively process the energetic dips in the masker but because the modulation produces a release from modulation masking; modulation masking is a contamination of the information in the speech envelope introduced by the intrinsic modulation of masking noise.

The intelligibility of speech in modulated noise from a different direction or modulated noises from several directions is relatively unexplored. Moreover, the studies that exist have come to contradictory or apparently contradictory conclusions.

Peissig and Kollmeier (1997) compared SRTs for speech masked by continuous noise and by competing speech. Between one and three interfering sources were located in virtual space around the listener. They found that SRM was greatly attenuated when multiple noise sources were used, particularly when interfering sources were in different hemifields. Interferers in multiple directions reduce the effectiveness of better-ear listening because neither ear is sheltered by the head from all of the interferers and because binaural unmasking is reduced when the overall masker complex is interaurally incoherent (Robinson and Jeffress, 1963; Culling *et al.*, 2004). However, Peissig and Kollmeier found that SRM was more robust when multiple speech interferers were used. They attributed this finding to an ability of the binaural system to alternately suppress interfering noise coming from different directions; because speech has a well-modulated amplitude envelope, considerable benefit could be gained by suppressing whichever of several voices

[a]Author to whom correspondence should be addressed. Electronic mail: CullingJ@cf.ac.uk

was more intense at a given moment. Peissig and Kollmeier discussed this potential effect in terms of the Equalization-Cancellation theory of binaural unmasking (Durlach, 1963, 1972) for which changes in the internal equalization delay will allow one or other sources to be canceled at different points in time. The same logic is applicable to better-ear listening. Listeners may "switch" back and forth, listening to one ear and then the other to follow the ear with the better SNR. We will term both such processes "binaural switching."

A major role of binaural switching in explaining the differences in SRM between speech and continuous-noise maskers was called into question by Hawley *et al.* (2004). They pointed out that speech and continuous-noise differ in more than just their amplitude envelopes. In similar experiments, they found that continuous and speech-modulated noises, which differed only in their amplitude envelopes, seemed to produce indistinguishable results, suggesting that speech envelopes facilitated very little binaural switching. Meanwhile, speech and reversed speech both produced considerably greater SRM, suggesting that properties other than modulation were critical in producing enhanced SRM with speech interferers. Much larger benefits of spatially separating speech interferers than either continuous or speech-modulated noises have also been found by Noble and Perrett (2002) and Jones and Litovsky (2011).

It is also interesting to contrast the concept of binaural switching with the effect of modulating a single masker on SRM. When only a single modulated noise is present, there is no opportunity for binaural switching. Rather than enhancing SRM, Goverts (2007) and George *et al.* (2012) found that the binaural unmasking component was *reduced* when a single interfering noise was modulated. They interpreted their results in terms of the need for "effective masker presence"; if there is no masker, then there cannot be any binaural unmasking, so every time there is a dip in the masker envelope, the SRM is reduced. Again, this reasoning could equally be applied to better-ear listening and so to SRM as a whole.

Masker modulation may thus have different effects upon the measured amount of SRM, depending on the number and spatial distribution of the masker sources.

When a second modulated masker is added, it both increases effective masker presence and, particularly if the masker is in a different hemifield from the first, offers the opportunity to employ binaural switching to gather speech information from different ears at different times and to suppress noises at different ITDs at different times. These three effects should increase SRM for the modulation- compared to the continuous-noise case.

Recently, Colburn *et al.* (2011) revisited the Hawley *et al.* data. They noted that although the effect of speech-modulation was small, the two-interferer case did display a 1.75 dB increase in SRM when using speech-modulated rather than continuous noise, suggesting that listeners had displayed some ability to enhance their performance through binaural switching. Moreover, they showed that a model of SRM that is capable of rapidly adapting to a changing masking configuration could predict these two-interferer data.

Hawley *et al.* performed all their experimental conditions with one, two, or three interferers, but, for the most part, presented and analyzed their data for each number of interferers separately. Given all the effects described above, one might expect that with one interferer, SRM would have been somewhat smaller for modulated than continuous noises, while for both two and three spatially distributed interferers, it would have been somewhat larger for modulated noise. Moreover, SRM differences in the two-interferer case would have been bigger than in the three-interferer case because the speech-modulated masker complex becomes less modulated the more independently modulated maskers are added. Figure 1 shows Hawley *et al.* SRM data as a function of the number of spatially distributed interferers; the data followed exactly the expected pattern, but the effects were small and perhaps not statistically significant. The present experiments set out first to produce a reliable demonstration of these effects (Experiment 1) and then to explore the individual roles of interaural time and level differences as well as the dependence on modulation rate (Experiment 2). The size of potential effects of modulation was maximized by using rectangularly modulated noise in simple virtual scenes with noise(s) at the optimum location(s) to promote SRM of a frontal speech source.

II. EXPERIMENT 1

Experiment 1 set out to simultaneously demonstrate the effects of effective masker presence and of binaural switching. To this end, interfering stimuli were created with either one or two maskers, which were either modulated or continuous. It was anticipated that a single modulated masker would produce less SRM than a continuous one but that two modulated maskers would produce more SRM than a continuous one. To make this cross-over interaction reliable, the masking stimuli were designed to produce large effects of both SRM and modulation. The maskers were placed on either side of the listener in virtual space at an azimuth of 105°. Peissig and Kollmeir (1997) found that 105° produced

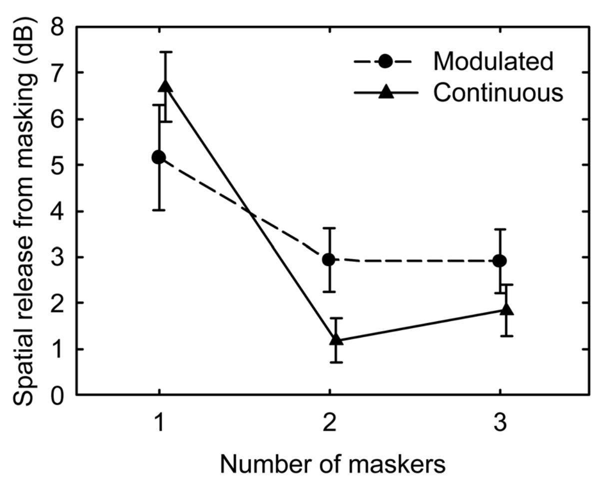

FIG. 1. Data replotted from the continuous–noise and speech-modulated-noise conditions of Hawley *et al.* (2004), showing spatial release from masking as a function of the number of spatially distributed maskers. Maskers were located at −30°, at −30° and 90°, or at −30°, 60°, and 90°. Error bars are 1 standard error of the mean.

the maximum effect of SRM for a frontal speech source. They observed an SRM of about 10 dB in this case. The maskers were modulated by a rectangular function with a 50% duty cycle; where two maskers were presented, they were modulated out-of-phase, such that the two together formed continuous noise when presented from the same location. Noble and Perrett (2002, Experiment 3) found that such alternating noises produced larger effects of switching than speech-modulated noise. For a single interferer of this type, the release from masking can exceed 20 dB (de Laat and Plomp, 1983). To avoid this effect overwhelming the results, the depth of modulation was limited to 20 dB.

A. Stimuli

The speech stimuli were the MIT recordings of voice DA speaking the IEEE sentences (Rothauser *et al.*, 1969). These stimuli are digitized at 20 kHz. The sentences were convolved with head-related impulse responses (HRIRs) from the KEMAR (Knowles Electronic Manikin for Acoustic Research) database collected by Gardner and Martin (1995) for 0° azimuth, 0° elevation (which had been resampled to 20 kHz for this purpose) to produced spatialized stereo stimuli.

Masking noises were also sampled at 20 kHz. They were generated by filtering two 5-s samples of white noise with a speech-shaped, 512-point FIR filter. This filter was designed to match the long-term excitation pattern (Moore and Glasberg, 1983) of the speech material. The resulting continuous speech-shaped noises were then modulated by each of two rectangular functions that alternated between 1 and 0.1 (i.e., a 20 dB modulation) with complementary duty cycles, at a rate of 10 Hz, to make modulated, speech-shaped noises. Both the modulated and the continuous speech-shaped noises were then spatialized by convolving them with the HRIRs for −105°, 0°, and +105° (0° elevation). Eight different interferer complexes were then created such that the number of maskers (one or two), the masker modulation (continuous or modulated) and the masker azimuths (0° or ±105°) were orthogonally manipulated. If there was only one masker, it was at +105°. If there were two maskers their modulation was complementary. The most complex case, with two modulated and spatially separated maskers is illustrated schematically in Fig. 2.

B. Procedure

Participants were seated in a single-walled IAC listening booth with a keyboard inside and an auxiliary monitor visible through the window. Stimuli were scaled and mixed digitally, converted to analog (Edirol UA 20) amplified (Project SE-II) and presented over headphones (Sennheiser HD650). The masking sound level was 65 dB (A) for two continuous noises in the frontal position. Masker level at each ear varied according to the number, location and modulation pattern of the masker(s). Participants made responses using the keyboard.

SRTs were measured using a computer-controlled technique similar to Plomp (1986) with the self-marking scheme introduced by Culling and Colburn (2000). A new implementation in MATLAB was used for the present experiments in

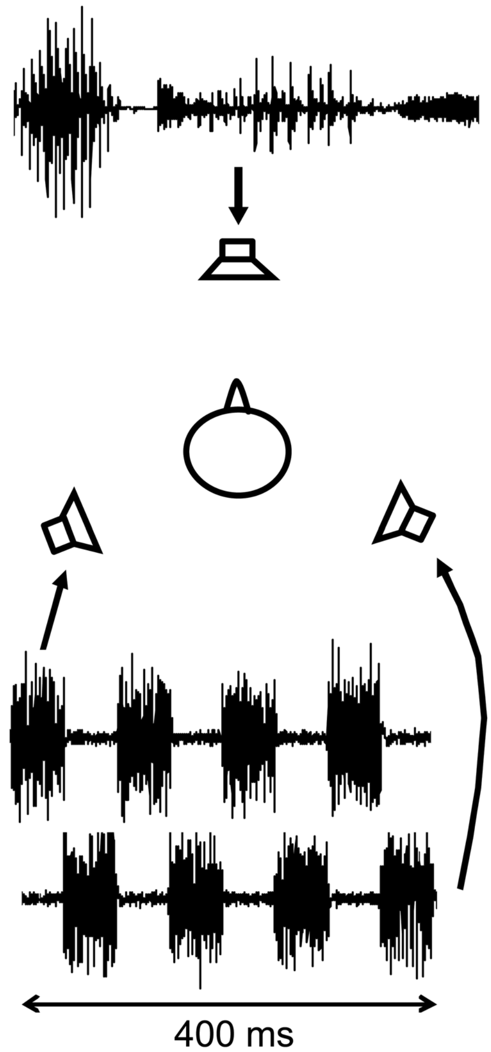

FIG. 2. Schematic illustration of the stimuli used in Experiment 1 when there were two modulated maskers at azimuths of ±105°. These maskers are similar also to those used in Experiment 2 except that the modulation was 100% and at various different rates in Experiment 2.

which transcripts of the first sentence were automatically monitored for accuracy.

Sentences were presented against the interferer complexes. The first sentence was initially presented at a very low SNR. Participants pressed the <enter> key on the keyboard to repeat the sentence with a higher SNR (the target sentence was increased in level by 4 dB), until they could hear enough speech to transcribe some words. The participant then attempted a transcript using the keyboard and auxiliary monitor and pressed <enter>. For the first sentence, the program performed a simple character-by-character, case-insensitive analysis of the words in the transcript to see whether at least two of them matched the keywords of the stored transcript. If less than two keywords matched, the response was neglected, and the first sentence was again increased in SNR without displaying the actual transcript. Otherwise, the participant proceeded with the self-marking procedure. Monitoring the first transcript in this way solves a persistent problem with the self-marking scheme. Participants occasionally attempt their initial transcript much too early, at a very low SNR. As a result, they achieve very little word recognition until the last few sentences, whereas the procedure assumes that last seven sentences are understood with approximately 50% intelligibility. The new procedure eliminated such events, reducing experimental noise and/or the need to discard contaminated data.

For each of the 10 sentences (including the first), the stored transcript was presented on the auxiliary monitor beneath the participants' transcript with the five keywords in capitals. The participant typed a digit between 0 and 5 to indicate how many of these five were correctly transcribed, and then the program passed on to the next sentence. Transcripts were now expected after a single presentation of each sentence. The computer increased the SNR by 2 dB if fewer than three words were correctly transcribed, and otherwise decreased the SNR by 2 dB. The new SNRs calculated *after* each of the last eight sentences were averaged to form the reported SRT.

Sixteen participants completed two practice SRT measurements using diotic continuous noise. They then did the eight experimental SRTs, one with each of the interferer complexes. Sentence materials were presented in the same order, but the experimental conditions (different interferer complexes) were in a pseudorandom order, which was rotated for each successive participant.

C. Results

The raw SRTs are presented in Table I. For statistical analysis, SRTs in the 0° conditions were subtracted from those in the corresponding ±105° conditions to derive measures of SRM. In this calculation, the effect of modulation on a single masker was factored out. This effect of dip listening was 9.7 dB for noise from the front. Like Fig. 1, Fig. 3 shows SRM as a function of the number of interfering noises for both continuous and modulated noise. For a single masker, SRM was smaller when that masker was modulated, but for two maskers, SRM was larger when the two maskers modulated out-of-phase with each other (such that the noise shifts from one side to the other) than when they were continuous.

A 2 × 2 analysis of variance for SRM showed a significant main effect of number of maskers [$F(1,15) = 79$, $p < 0.001$] and a significant interaction between number of maskers and masker modulation [$F(1,15) = 17.9$, $p < 0.001$]. From Fig. 2 it is clear that this is a cross-over interaction, consistent with the hypothesis.

D. Discussion

Experiment 1 set out to generate a reliable demonstration of both the effects of effective masker presence and of binaural switching within the same experiment. The combination of these effects can be seen from the significant cross-over interaction; SRM was greater for continuous than for modulated noise when there was a single masker (effective masker presence), but SRM was greater for modulated than continuous noise for two simultaneous maskers (binaural switching). The switching effect was also observed by Noble and Perrett (2002) using a similar design but much slower alternation (loudspeaker presentation of noises switching between ±90° every 250 ms). Models of SRM thus need to

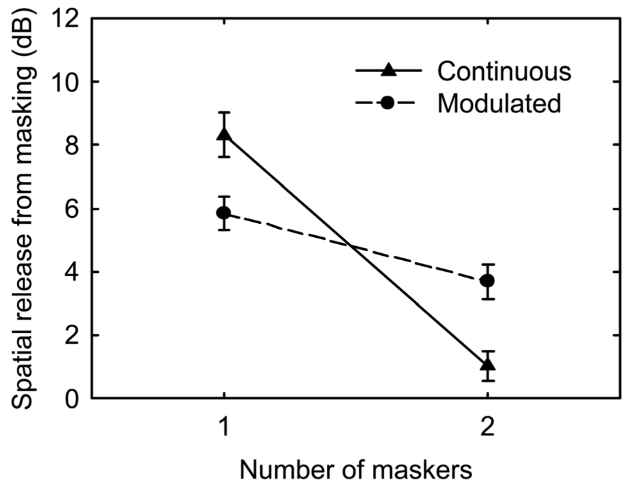

FIG. 3. Data from Experiment 1. Spatial release from masking for one or two spatially distributed maskers, which were either continuous or block-modulated speech-shaped noise. Maskers were located at 105° or at 105° and 255°. Error bars are 1 standard error of the mean.

incorporate such effects to make accurate predictions for multiple modulated interferers.

The SRM was maximal at 8.3 dB for a single continuous noise source at +105°. This value is comparable to the SRM of 9.8 dB observed by Peissig and Kollmeier (1997) and the prediction of 10 dB made by Jelfs *et al.* (2011) for speech at 0° and noise at +105°. When the same masker was rectangularly modulated at 10 Hz, to a depth of 20 dB, the reduction in effective masker presence reduced the SRM to 5.8 dB. It should be remembered, however, that due to dip listening, the underlying SRTs for a single modulated noise were substantially lower than those for the single continuous noise; it is only the further improvement with spatial separation (SRM) that is reduced by the modulation (see Table I).

When two maskers were presented simultaneously, SRM was always reduced. SRM was minimal at 1.0 dB when continuous interferers were presented from both −105° and +105°, which is quite low compared to Peissig and Kollmeier's measurement of 4.1 dB and predictions from the Jelfs *et al.* model of 3.2 dB.[2] SRM was relatively robust (4 dB) when the two noises were modulated out of phase with each other. In this situation, the noise effectively shifts back and forth to left and right at the modulation rate, and the listener has the opportunity to use binaural switching to maintain SRM. However, if binaural switching worked perfectly, one might expect SRM to be almost unaffected by the addition of the second modulated noise because at any given

TABLE I. Speech reception thresholds from Experiment 1 calculated as the ratio of speech level at source to the combined noise level at source (i.e., prior to spatialization).

	Noises			
	1 noise		2 noises	
Locations	Continuous	Modulated	Continuous	Modulated
Co-located	−4.5	−11.4	−4.6	−4.4
Separated	−12.8	−17.2	−5.7	−8.1

moment one masker is always attenuated by 20 dB and should play relatively little role in determining SRT. SRM was in fact reduced by 1.5 dB for such pairs of modulated maskers compared to the SRM observed for a single modulated masker. Moreover, the magnitude of SRM for two modulated maskers (4 dB) seems to be only marginally higher than that predicted for two continuous maskers (3.2 dB). These observations together suggest that there is some limiting factor.

As noted earlier, SRM can be considered to have two components, better-ear listening and binaural unmasking (Zurek, 1993; Lavander and Culling, 2010; Jelfs *et al.*, 2011). Each of these components could be limited by the interaural alternation rate when two maskers in opposite hemifields are modulated out of phase because the binaural system is known to adapt rather slowly to changing interaural stimulus parameters (e.g., Grantham and Wightman, 1978). Peissig and Kollmeier considered the problem from the perspective of Equalization-Cancellation theory and pointed out that sluggishness in binaural unmasking (Grantham and Wightman, 1979; Kollmeier and Gilkey, 1990; Culling and Summerfield, 1998) would predict a decline in SRM with increasing modulation rate. Similarly, there may be some limit to the speed with which the binaural system can switch between ears to follow the most favorable SNR.

III. EXPERIMENT 2

Experiment 2 was designed to measure the potential influence of binaural sluggishness on listeners' ability to perform binaural switching. Binaural unmasking relies on differences between target and masker in ITDs, while better-ear listening relies on differences in interaural level differences (ILDs). Some previous work has suggested that processes based on ITDs and those based on ILDs may be subject to different time constants. For instance, Grantham and Wightman (1978) measured the detection of modulation in ITD, while Grantham (1984) measured similar detection of modulations in ILD. Grantham (1984, p. 71) noted that the binaural system appeared to follow the ILD modulations "more efficiently," meaning that sensitivity to modulations was maintained at a higher modulation rates in the case of ILDs. In the present experiment, therefore, we measured the individual and combined effects of ILDs and ITDs on binaural switching as a function of modulation rate, using similar techniques to Bronkhorst and Plomp (1988) and Culling *et al.* (2004).

Because the focus was exclusively on binaural switching, all test conditions involved two maskers at $\pm 105°$ that were rectangularly modulated in alternation. Because there was no single-masker condition, it was no longer necessary to limit the depth of modulation. All conditions thus required a common control condition in which SRM was not possible; the control condition had a single continuous masking noise at 0°. Different test conditions included different combinations of interaural differences (ILDs, ITDs or both) and different rates of rectangular modulation (1, 2, 5, 10, and 20 Hz). These rates were selected partly because they bracketed those used in Experiment 1 and partly because they covered the range over which the effects of sluggishness might be expected to vary. Experiment 1 produced a switching-based effect of SRM using a 10 Hz modulation rate, but this effect was substantially smaller than the SRM observed with a single continuous masker. If this deficit in performance is attributable to sluggishness, then one might expect SRM to be larger at lower rates, such as 2 or 5 Hz, and smaller still at even higher rates, such as 20 Hz. The 100-ms modulation cycle used in Experiment 1 was also comparable to the time constants of 40-200 ms measured in various studies (Kollmeier and Gilkey, 1990; Culling and Summerfield, 1998; Akeroyd and Summerfield, 1999; Kolarik and Culling, 2009), suggesting that modulation rates in this range would be expected to be influenced by sluggishness.

A. Stimuli

Speech stimuli were similar to those of Experiment 1. The speech-shaped masking noises were 100% modulated with a 50% duty cycle at rates of 1, 2, 5, 10, and 20 Hz. Two noises were created at each modulation rate with complementary duty cycles.

Three types of HRIR were prepared. They were either drawn directly from the KEMAR database as in Experiment 1 or were first manipulated to remove one interaural disparity or the other. For this purpose, the amplitude and phase spectra of the HRIRs for $-105°$, 0°, and $+105°$ were extracted. To isolate ILDs, the amplitude spectra for $-105°$ or $+105°$ were combined with the phase spectra for 0° and converted back into impulse responses using inverse Fourier transformation. To isolate ITDs, the phase spectra for $-105°$ or $+105°$ were combined with the amplitude spectra for 0° and converted back into impulse responses using inverse Fourier transformation. As in Experiment 1, all HRIRs were resampled to 20 kHz sampling rate.

The resulting ITD + ILD, ITD-only, and ILD-only HRIRs were convolved with the modulated speech-shaped noises similar to those of Experiment 1 but with 100% modulation. One from each pair of noises with complementary duty cycles was convolved with the HRIRs for -105° and the other with the HRIRs for $+105°$. These pairs were then added together to make 15 different two-masker complexes (5 modulation rates $\times$ 3 HRIR types). A 16th masker type was prepared with continuous speech-shaped noise convolved with the HRIR for 0°. Speech stimuli were convolved with the HRIR for 0°.

B. Procedure

Sixteen participants attended a single 90-min experimental session. SRTs were measured using a similar method to Experiment 1 with two practice measurements and 16 experimental measurements. The 16 experimental measurements used each of the different masker stimuli described above in a pseudorandom order that was rotated for each successive participant.

C. Results

SRTs using each of the 15 two-masker complexes were subtracted from those from the same participants for continuous noise at 0° to yield measures of SRM. The mean SRT at 0° was -4.4 dB relative to the level of the combined masking noises. Figure 4 shows the mean SRMs as a function of modulation rate for each of the three types of HRIR. There was a decline in SRM with increasing modulation rate, which appears to asymptote toward 20 Hz at about 4 dB. For comparison, Fig. 4 also shows the equivalent predictions of the Jelfs *et al*. model based on the HRIRs used in the experiment for a continuous noise at +105° (open symbols), and the corresponding data point from Experiment 1 for the ILD + ITD case.

A two-way analysis of variance (3 HRIR types × 5 modulation rates) confirmed significant effects of modulation rate [$F(4,60) = 18$, $p < 0.001$] and HRIR type [$F(2,30) = 73$, $p < 0.001$] as well as an interaction between the two [$F(8,120) = 4.7$, $p < 0.001$].

D. Discussion

SRM declined with increasing modulation rate for modulated ILDs, modulated ITDs, and their combination, suggesting that the binaural switching process is rather sluggish. Even at a rate of 1 Hz, SRM was substantially lower than when measured for a single continuous interfering source at 105° in Experiment 1 (filled upright triangle at 0 Hz modulation rate in Fig. 3), or as predicted by the Jelfs *et al*. model (open upright triangles). The model also provides predictions for static noises using ILDs and ITDs in isolation (open symbols in Fig. 4), which suggest that the effects of the individual components of SRM are each reduced at the 1-Hz modulation rate.

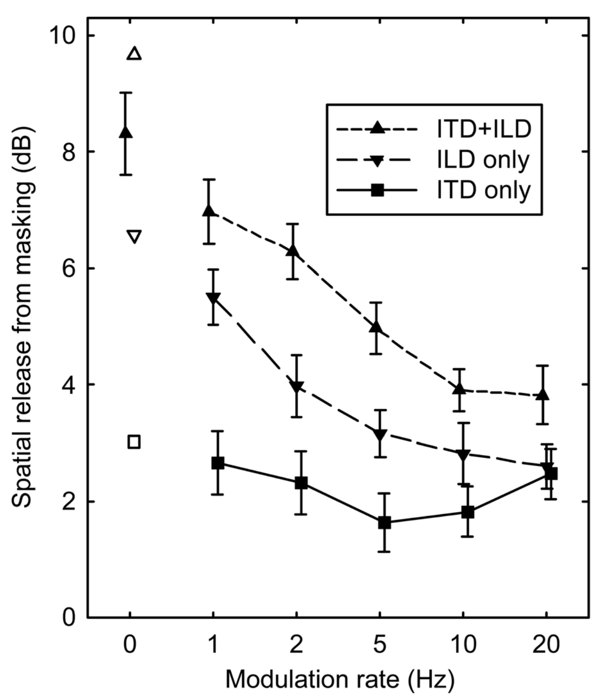

FIG. 4. Data from Experiment 2. Spatial release from masking as a function of modulation rate for maskers at 105° and 255° and for simulated locations generated using both ILDs and ITDs (closed upright triangles) and as well as each cue in isolation (inverted triangles and squares). Error bars are 1 standard error of the mean. For comparison predictions of the Jelfs *et al*. (2011) model for continuous a continuous noise at 105° (open symbols) and equivalent data from Experiment 1 for a continuous noise (isolated close upright triangle) are shown at 0 Hz modulation rate.

The significant interaction reflects, to some extent, the fact that the different cues and their combination produce different magnitudes of effect; if these different effects were all to asymptote to zero with increasing modulation rate, then one would see such a statistical interaction. It is, therefore, unclear whether there is a difference between binaural cues in the steepness of decline in SRM with modulation rate. Figure 4 does not suggest, however, that there is a large difference in the sluggishness between ILDs and ITDs in this context.

Measures of sluggishness are often subject to confounding variables, such as the introduction of interaural incoherence [see discussion of Grantham and Wightman (1978), Grantham (1984), and Kolarik and Culling (2009)]. The present study does not appear to suffer from this problem. Moreover, we are not aware of any previous study in which sluggishness of ILD processing has affected a spatial unmasking process. Because speech intelligibility relies on the integration of information across time and frequency, the present method offers a means of accessing the auditory system's ability to cope with dynamic binaural maskers and to make sensible comparisons between its effects on the processing of ILDs and ITDs.

The effect observed here of sluggishness for modulated ITDs may be the consequence of a limited processing resolution, as embodied in the concept of a binaural temporal window (Kollmeier and Gilkey, 1990; Culling and Summerfield, 1998; Akeroyd and Summerfield, 1999). In this scheme, it is generally supposed that the limit in temporal resolution precedes the selection of an appropriate cancellation delay (Kohlrausch, 1990; Culling and Colburn, 2000) rather than suggesting that it requires time for the binaural system to change its cancellation delay (Yost, 1985). However, the concept of an early limit on temporal resolution, represented by a sliding temporal window, does not seem adequate to explain our results with alternating ILDs. Because listeners have much shorter monaural temporal windows (Plack and Moore, 1990), which should be adequate to resolve information in the temporal dips at each ear, it seems necessary to place the limit in temporal resolution at the level of selection of appropriate information from the two ears, i.e., how quickly the auditory system can switch between selecting one ear and selecting the other.

Neither component effect declines toward zero SRM, but this should be expected. The effect of ILDs should converge on the level achievable by simply listening against the modulated noise at one ear (a level which would depend on modulation rate). The effect of ITDs will be sustained by listeners' ability to cope with two simultaneous maskers at different ITDs (Culling *et al*., 2004).

Consistent with the assumption made by the Lavandier and Culling (2010) and Jelfs *et al*. (2011) models of spatial unmasking, SRM derived from isolated ILDs and ITDs was approximately additive; the sum of the two mean SRMs derived from the individual cues (ILD-only) and (ITD-only) exceeded that for the combined cues (ILD + ITD) by an average of 0.6 dB (s.d. = 0.6 dB).

IV. GENERAL DISCUSSION

The results of Experiment 1 simultaneously confirm the expected effects of effective masker presence and of binaural switching. Listeners displayed a reduced benefit from spatial separation of target and masking sources when the masking source was strongly modulated, reducing its presence. On the other hand, they displayed an improved benefit from spatial separation of a target source from two masking sources when those two sources were modulated (given that the modulation functions were different). However, Experiment 2 showed that this binaural switching is a fairly sluggish process, requiring modulation rates below about 5 Hz for substantial effects to be observed. This finding suggests that the increased SRM observed in Experiment 1 at a 10 Hz modulation rate may have been partly generated by an unusually low estimate of SRM (1 dB) for continuous noises on both sides (±105°).

Our conclusion that binaural switching is very sluggish contrasts with that drawn by Brungart and Iyer (2012). Using a similar listening situation, but with two interfering voices at ±90°, they found that stimuli pre-processed to present the most favorable SNR consistently to the same ear gave similar intelligibility to the unprocessed stimuli. They calculated the modal glimpse length available to each ear in the unprocessed case to be about 100 ms, suggesting interaural switching at least 5 Hz was needed in the unprocessed case. It is possible that the failure of the processing to yield greater improvement in intelligibility can be attributed to processing artifacts.

The dominant modulation frequency in speech is 3–4 Hz, corresponding to the rate at which syllables are uttered (Drullman *et al.*, 1993). From the results of Experiment 2, binaural switching would have limited effects at this rate of modulation. The modulation of speech is also more graded than the square-wave modulation used in the present experiments, which would mean that improvements in SNR at each ear would be more modest for speech than for the square-wave modulation used in our experiments. Noble and Perrett (2002) observed about 2.5 dB of SRM for 2-Hz square-wave modulated noise (Experiment 2) but less that 1 dB for speech modulated maskers. These two factors together suggest that the practical importance of binaural switching may be quite limited, and this interpretation is supported by the rather marginal effect seen in Fig. 1 where speech-modulated noises were used. Hawley *et al.* results (Fig. 1) suggest that the effect is less than 2 dB for two speech interferers.

It seems that Hawley *et al.* were at least partly correct in suggesting that binaural switching was not responsible for the substantial differences in the robustness of SRM between speech and noise interferers observed by Peissig and Kollmeier. Instead, Hawley *et al.* found evidence for enhanced SRM for both speech and reversed speech but not for continuous or speech-modulated noise. It is tempting to suppose that the enhanced SRM that they found may be related to release of informational masking, but this interpretation is difficult to sustain when considering the absolute SRTs; SRTs were consistently higher for speech than for the other three interferer types, suggesting that only speech and not reversed speech was generating substantial informational masking. The mechanisms underlying this effect therefore remain poorly understood.

It is important to include the effects of effective masker presence and binaural switching in future models of SRM. However, the present data indicate that a strict rate limitation should be included in the switching capacity of such a model. These data also provide a useful dataset against which the predictions of such a model could be evaluated.

V. CONCLUSIONS

Modulation of masking sources has complex but subtle effects on SRM. For a single interfering source, the reduction in effective masker presence produces a corresponding reduction in release from that masking. For multiple interfering sources, binaural switching, based on interaural differences in both level and timing can increase SRM, but this process is severely rate limited such that the modulation rate of a speech interferer will reduce the magnitude to a decibel or two.

ACKNOWLEDGMENT

We would like to thank the two reviewers for many helpful comments in the development of this paper.

[1]Good predictions were made for block-modulated noise with various duty cycles and for speech-modulated noise, but the effect of modulation rate for sinusoidally modulated noise was not predicted.

[2]A release from masking is predicted in this case mainly because binaural unmasking is fairly robust for multiple maskers in anechoic conditions (Culling *et al.*, 2004) There is also a small predicted better-ear listening effect of 0.2 dB that may be due to reductions in overall masker level.

Akeroyd, M. A., and Summerfield, A. Q. (**1999**). "A binaural analog of gap detection," J. Acoust. Soc. Am. **105**, 2807–2820.

ANSI (**1997**). ANSI S3.5-1997, "Methods for calculation of the speech intelligibility index" (American National Standards Institute, New York).

Beutelmann, R., and Brand, T. (**2006**). "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **120**, 331–342.

Beutelmann, R., Brand, T., and Kollmeier, B. (**2010**). "Revision, extension and evaluation of a binaural speech intelligibility model," J. Acoust. Soc. Am. **127**, 2479–2497.

Bronkhorst, A. W., and Plomp, R. (**1988**). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," J. Acoust. Soc. Am. **83**, 1508–1516.

Brungart, D. S., and Iyer, N. (**2012**). "Better-ear glimpsing efficiency with symmetrically placed interfering talkers," J. Acoust. Soc. Am. **132**, 2545–2556.

Colburn, H. S., Wan, R., and Durlach, N. I. (**2011**). "Predicting speech intelligibility with multiple interfering sources," in *Forum Acusticum* (Aalborg).

Culling, J. F. (**2007**). "Evidence specifically favoring the Equalization-Cancellation theory of binaural unmasking," J. Acoust. Soc. Am. **122**, 2803–2813.

Culling, J. F., and Colburn, H. S. (**2000**). "Binaural sluggishness in the perception of tone sequences and speech in noise," J. Acoust. Soc. Am. **107**, 517–527.

Culling, J. F., Hawley, M. L., and Litovsky, R. Y. (**2004**). "The role of head-induced interaural time and level differences in the speech reception

threshold for multiple interfering sound sources," J. Acoust. Soc. Am. **116**, 1057–1065.
Culling, J. F., and Summerfield, Q. (**1998**). "Measurements of the binaural temporal window using a detection task," J. Acoust. Soc. Am. **103**, 3540–3553.
de Laat, J. A. P. M., and Plomp, R. (**1983**). "The reception threshold of interrupted speech" in *Hearing: Phsyiological Bases and Psychophysics*, edited by R. Kinke and R. Hartman (Springer, Berlin) pp. 359–363.
Drullman, R., Festen, J. M., and Plomp, R. (**1993**). "Effect of temporal envelope smearing on speech reception," J. Acoust. Soc. Am. **95**, 1053–1064.
Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking level differences," J. Acoust. Soc. Am. **35**, 1206–1218.
Durlach, N. I. (**1972**). "Binaural signal detection: Equalization and cancellation theory" in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York). Vol. II, pp. 369–462.
Festen, J. M. (**1987**). "Speech-perception threshold in a fluctuating background sound and its possible relation to temporal resolution" in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Martinus Nijhoff, Dordrecht), pp. 461–466.
Gardner, W. G., and Martin, K. D. (**1995**). "HRTF measurements of a KEMAR," J. Acoust. Soc. Am. **97**, 3907–3908.
George, E. J., Festen, J. M., and Goverts, S. T. (**2012**). "Effects of reverberation and masker fluctuations on binaural unmasking of speech," J. Acoust. Soc. Am. **132**, 1581–1591.
Goverts, S. T. (**2007**). "The influence of masker type on the binaural intelligibility level difference" in *Hearing—From Sensory Processing to Perception*, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauerman, and J. Verhey (Springer, Berlin) pp. 551–558.
Grantham, D. W. (**1984**). "Discrimination of dynamic interaural intensity differences," J. Acoust. Soc. Am. **76**, 71–76.
Grantham, D. W., and Wightman, F. L. (**1978**). "Detectability of varying interaural temporal differences," J. Acoust. Soc. Am. **63**, 511–523.
Grantham, D. W., and Wightman, F. L. (**1979**). "Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation," J. Acoust. Soc. Am. **65**, 1509–1517.
Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (**2004**). "The benefit of binaural hearing in a cocktail party: Effect of location and type of masker," J. Acoust. Soc. Am. **115**, 833–843.
Jelfs, S., Lavandier, M., and Culling, J. F. (**2011**). "Revision and validation of a binaural model for speech intelligibility in noise," Hear. Res. **275**, 96–104.
Jones, G. L., and Litovsky, R. Y. (**2011**). "A cocktail party model of spatial release from masking by both noise and speech interferers," J. Acoust. Soc. Am. **130**, 1463–1474.
Kohlrausch, A. (**1990**). "Binaural masking experiments using noise maskers with frequency-dependent interaural phase differences. II: Influence of frequency and interaural-phase uncertainty," J. Acoust. Soc. Am. **88**, 1749–1756.
Kolarik, A. J., and Culling, J. F. (**2009**). "Measurement of the binaural temporal window using a lateralization task," Hear. Res. **248**, 60–68.
Kollmeier, B., and Gilkey, R. H. (**1990**). "Binaural forward and backward masking: Evidence for sluggishness in binaural detection," J. Acoust. Soc. Am. **87**, 1709–1719.
Lavandier, M., and Culling, J. F. (**2010**). "Prediction of binaural speech intelligibility against noise in rooms," J. Acoust. Soc. Am. **127**, 387–399.
Miller, G. A., and Licklider, J. C. R. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**, 167–173.
Moore, B. C. J., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.
Noble, W., and Perrett, S. (**2002**). "Hearing speech against spatially separate competing speech versus competing noise," Percept. Psychophys. **64**, 1325–1336.
Peissig, J., and Kollmeier, B. (**1997**). "Directivity of binaural noise reduction in spatial multiple-source arrangements for normal and impaired listeners," J. Acoust. Soc. Am. **101**, 1660–1670.
Plack, C. J., and Moore, B. C. J. (**1990**). "Temporal window shape as a function of frequency and level," J. Acoust. Soc. Am. **87**, 2178–2187.
Plomp, R. (**1986**). "A signal-to-noise ratio method for the speech-reception SRT of the hearing impaired," J. Speech Hear. Res. **29**, 146–154.
Rhebergen, K. S., and Versfeld, N. J. (**2005**). "A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," J. Acoust. Soc. Am. **117**, 2181–2192.
Robinson, D. E., and Jeffress, L. A. (**1963**). "Effect of varying the interaural noise correlation on the detectability of tonal signals," J. Acoust. Soc. Am. **35**(12), 1947–1952.
Rothauser, E. H., Chapman, W. D., Guttman, N., Nordby, K. S., Silbiger, H. R., Urbanek, G. E., and Weinstock, M. (**1969**). "I.E.E.E. recommended practice for speech quality measurements," IEEE Trans. Aud. Electroacoust. **17**, 227–246.
Stone, M. A., Füllgraber, C., and Moore, B. C. J. (**2012**). "Notionally steady background noise acts primarily as a modulation masker of speech," J. Acoust. Soc. Am. **132**, 317–326.
Wan, R., Durlach, N. I., and Colburn, H. S. (**2010**). "Application of an extended equalization-cancellation model to speech intelligibility with spatially distributed maskers," J. Acoust. Soc. Am. **128**, 3678–3690.
Yost. W. A. (**1985**). "Prior stimulation and the masking level difference," J. Acoust. Soc. Am. **78**, 901–907.
Zurek, P. M. (**1993**). "Binaural advantages and directional effects in speech intelligibility" in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. Studebaker and I. Hockberg (Allyn and Bacon, Needham Heights, MA) pp. 255–276.