

Rhythmic masking release: Contribution of cues for perceptual organization to the cross-spectral fusion of concurrent narrow-band noises^{a)}

Martine Turgeon,^{b)} Albert S. Bregman, and Pierre A. Ahad

Psychology Department, McGill University, 1205 Dr. Penfield Avenue, Montreal, Quebec, Canada

(Received 9 March 2000; revised 27 October 2000; accepted 29 December 2001)

The contribution of temporal asynchrony, spatial separation, and frequency separation to the cross-spectral fusion of temporally contiguous brief narrow-band noise bursts was studied using the Rhythmic Masking Release paradigm (RMR). RMR involves the discrimination of one of two possible rhythms, despite perceptual masking of the rhythm by an irregular sequence of sounds identical to the rhythmic bursts, interleaved among them. The release of the rhythm from masking can be induced by causing the fusion of the irregular interfering sounds with concurrent “flanking” sounds situated in different frequency regions. The accuracy and the rated clarity of the identified rhythm in a 2-AFC procedure were employed to estimate the degree of fusion of the interfering sounds with flanking sounds. The results suggest that while synchrony fully fuses short-duration noise bursts across frequency and across space (i.e., across ears and loudspeakers), an asynchrony of 20–40 ms produces no fusion. Intermediate asynchronies of 10–20 ms produce partial fusion, where the presence of other cues is critical for unambiguous grouping. Though frequency and spatial separation reduced fusion, neither of these manipulations was sufficient to abolish it. For the parameters varied in this study, stimulus onset asynchrony was the dominant cue determining fusion, but there were additive effects of the other cues. Temporal synchrony appears to be critical in determining whether brief sounds with abrupt onsets and offsets are heard as one event or more than one. © 2002 Acoustical Society of America. [DOI: 10.1121/1.1453450]

PACS numbers: 43.66.Dc, 43.66.Lj, 43.66.Mk, 43.66.Rq [DWG]

I. INTRODUCTION

A. Multiple cues in sound-source determination

Many scientists have studied the question of how the auditory system parses the acoustic signal so as to provide the animal with a useful perceptual description of the activity of individual sound sources (Bregman, 1990, 1993; Darwin and Carlyon, 1995; Hartmann, 1988; Moore, 1989; Yost, 1991). Bregman (1990) has proposed that the auditory parsing process is governed by ecologically valid heuristics that have evolved to exploit the acoustical properties of causally related sound-producing events. There has been converging empirical evidence that the auditory system is built to extract these regularities from the acoustic signal for the purpose of sound-source determination. This has been reviewed by several researchers (Bregman, 1990; Darwin and Carlyon, 1995; Yost and Sheft, 1993; Yost, 1991).

No auditory grouping¹ cue operates in isolation; rather cues act together; sometimes reinforcing each other, and sometimes competing with each other to provide the groupings of components upon which the most valid perceptual description of the acoustic signal can be built. That cues can have combined effects is a recognized fact and stimulated empirical work some 20 years ago (Bregman, 1978; Breg-

man and Pinker, 1978; Dannenbring and Bregman, 1978; Steiger and Bregman, 1982); until recently, there has been relatively little subsequent work done on auditory organization in the presence of multiple cues. One of the main goals of the present study was to further explore the perceptual outcome when factors known to either promote the segregation or the fusion² of complex sounds, act together.

Apart from the recognition of multiple cues in sound-source determination, there has been a growing recognition of the importance of cross-spectral analysis (Yost and Sheft, 1993). A variety of paradigms have been employed to investigate the cross-spectral integration¹ of acoustical information: profile analysis (Green, 1988), modulation detection interference or MDI (Hall and Grose, 1991; Yost *et al.*, 1989), comodulation masking release or CMR (Grose and Hall, 1993; Hall *et al.*, 1984), comodulation detection difference or CDD (McFadden and Wright, 1990), and more recently, comodulation masking protection or CMP (Gordon, 1997). All of those paradigms are based upon an analysis of energy across frequency channels, though they differ as to the task, stimuli and measurements used to explore cross-spectral integration. For instance, while in CMR the detection of a sinusoidal target signal masked by a modulated noise within the same frequency band is improved by the presence of comodulated flanker noises situated in different frequency bands, in MDI the discrimination of the depth of modulation of such a target can be impaired by the presence of comodulated sinusoidal maskers of different frequencies. Despite these methodological differences, these paradigms provide converging evidence that the auditory system is sensitive to

^{a)}This research was presented as part of the first author's Ph.D. thesis to the Psychology Department of McGill University.

^{b)}Reprints are available from Martine Turgeon at “Behavioural Brain Sciences Centre, School of Psychology, The University of Birmingham, Edgbaston, Birmingham B15 2TT, UK,” where she is currently affiliated. Electronic mail: M.Turgeon@Bham.ac.uk

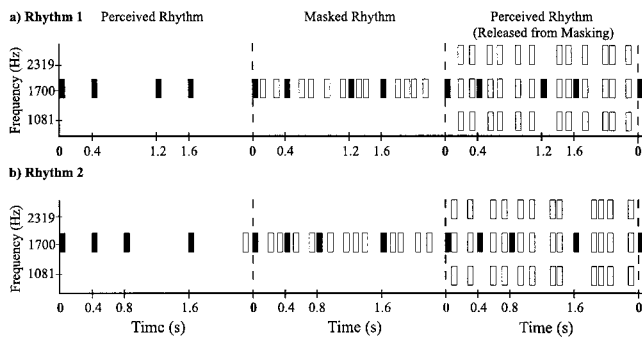


FIG. 1. Temporal structures of sequences in rhythmic masking release (RMR). Rhythms 1 and 2 consist of 3.5 replications of the succession of intervals shown at the left of panel (a) and (b), respectively. They are composed of different temporal arrangements of 48-ms noise bursts separated by two possible intervals: 384 and 768 ms. To camouflage perceptually the rhythm, two irregular “maskers” are added in the 384-ms interval and four in the 768-ms one [see middle of panels (a) and (b)]. The rhythm is masked because no acoustic property distinguishes the regular from the irregular sounds. The rhythm can be released from masking when “flankers” of different frequencies are added simultaneously to the maskers, as shown on the right of panels (a) and (b). Hearing the rhythm depends on the fusion of the irregular maskers and flankers.

across-frequency correlations in the time-varying patterns of intensity.

B. RMR to study cross-spectral fusion in the context of multiple cues

The present study uses the rhythmic masking release (RMR) paradigm (Bregman and Ahad, 1996, Demonstration 22; Turgeon and Bregman, 1997) to look directly at the link between cross-spectral integration and perceptual fusion. In RMR, perceptual fusion depends on the use of relations among the components of the signal in different frequency regions, such as simultaneous onsets and offsets. Figure 1 schematizes the temporal structure of the stimuli used in the present study. Similar stimuli were used in a preliminary RMR experiment (Turgeon, 1999). The results of that experiment are useful to introduce the RMR paradigm. They also provide some predictions as to what should be perceived under different conditions of the present study. When a regular sequence of narrowband noise bursts is played in isolation, a simple rhythm is heard. Such rhythms are perceived upon repeating the successions of short and long intervals shown in the left of panels (a) and (b) of Fig. 1. While alternating the short and long intervals shown in the top pattern (a) evokes a rhythm with pairs of bursts (Rhythm 1 in the present study); cycling the succession of short, long, long, and short intervals shown in the bottom pattern (b) evokes a rhythm with triplets of bursts alternating with a single burst (Rhythm 2 in the present study). In both examples, sounds that are closer together in time perceptually group together (Handel, 1989).

If an irregular sequence of identical sounds is intermingled among those of the regular one, the rhythm is no longer heard (white bars of the same frequency as the dark bars). This is because no acoustic property distinguishes the regular bursts from the irregular ones. We refer to the camouflaging bursts as “maskers;” while they do not mask the regular bursts, they do mask their sequential organization,

that is, the rhythm.³ Together, the rhythmic bursts and the maskers form the “masked-rhythm sequence” (dark and white bars in the middle portion of Fig. 1). In the rightmost portion, narrowband noise “flankers” are added. These are synchronous with the irregular maskers, but located in other frequency regions (white bars at the top and bottom of the masked rhythm sequence). The prior RMR study has shown that this causes the rhythm to be “released from masking” (Turgeon, 1999). The release was explained by the fusion of the maskers and flankers that have simultaneous onsets. The emergent perceptual properties of the masker-flanker complexes (e.g., a global timbre different from that of each rhythmic pulse) allow the listener to distinguish these irregularly spaced bursts from the regularly-spaced ones. Hence, the accurate perception of the rhythm is contingent upon the fusion of the two irregular sequences of maskers and flankers into a single sequence of masker-flanker complexes.

II. EXPERIMENT 1. TEMPORAL LIMITS AND RELATIVE CONTRIBUTION OF CUES TO THE CROSS-SPECTRAL FUSION OF NOISE BURSTS PRESENTED BINAURALLY

Experiment 1 explored the contribution of four acoustical properties to diotic and dichotic fusion: temporal asynchrony (Dannenbring and Bregman, 1978; Darwin and Ciocca, 1992), amplitude modulation (Bregman *et al.*, 1985; Grose and Hall, 1993) frequency separation (Brochard *et al.*, 1999; Turgeon, 1994) and dichotic presentation (Hukin and Darwin, 1995; Kidd *et al.*, 1994). There is evidence that each of these properties influences the cross-spectral integration of information in a number of phenomena: (i) Temporal asynchrony has been shown to affect MDI by Hall and Grose (1991), binaural MDI by Sheft and Yost (1997), CMR by Grose and Hall (1993) and McFadden (1986), CMP by Gordon (1997), and localization by Woods and Colburn (1992). (ii) The correlation of envelope modulation across frequency has been related to CMR by Grose and Hall (1993), to dichotic CMR by Schooneveldt and Moore (1987), to MDI by Yost *et al.* (1989), and to binaural MDI by Sheft and Yost (1997). (iii) The frequency separation between the signal and the flanking bands decreases CMR (Schooneveldt and Moore⁴). (iv) Similarly, the contralateral presentation of the flanking bands relative to the signal decreases CMR (Schooneveldt and Moore, 1987) and reduces MDI slightly (Sheft and Yost, 1997).

In this study, the asynchrony of onset and offset between pairs of narrow-band flanker bursts symmetrically placed relative to each masker burst was manipulated. Because the maskers and flankers had the same duration, the onset asynchrony was equal to the offset asynchrony; hence only the magnitude of the stimulus onset asynchrony (SOA) will be mentioned. The ecological validity of temporal synchrony for sound-source determination combined with the converging evidence that it is a very powerful grouping cue (Darwin and Carlyon, 1995), including a prior RMR experiment (Turgeon, 1999), led us to the following hypotheses: (i) Synchronous maskers and flankers should fully fuse to yield RMR independently of their frequency separation (ΔF), whether or not the flankers are presented in the same or contralateral

ear as the maskers; (ii) SOA should significantly decrease fusion for both the diotic and dichotic presentation of concurrent maskers and flankers; and (iii) SOA, ΔF and contralateral presentation should reinforce each other in favoring the segregation (i.e., diminishing the fusion) of the concurrent maskers and flankers.

Another important goal was to estimate the temporal asynchrony required to abolish the fusion of concurrent sounds situated in different frequency regions. For convenience, we refer to such an asynchrony as “SOA threshold,” though we do not suggest that it applies to synchrony per se; it is rather an “event-segregation threshold,” that is, the SOA necessary to perceive brief sounds close together in time, as separate events. A last objective was to compare the cross-spectral fusion resulting from correlated amplitude fluctuations at different temporal scales, namely the slow amplitude changes at the macro scale of the whole sequence (i.e., onsets and offsets) with the faster ones at the micro scale within each sequential component (AM).

Grose and Hall (1993) wanted to know whether a correlated pattern of AM was sufficient to induce CMR. Although it was shown to induce CMR when the onsets of correlated masking and flanking bands were simultaneous, the CMR was considerably decreased when they were asynchronous; in fact, a 50-ms SOA between the on-signal band and the flanking bands completely abolished CMR. Note that the asynchronous bands were comodulated during their period of overlap. These results are consistent with those of McFadden (1986) who found that SOAs between 3-to-15 ms abolished CMR. This suggests that the effect of a common AM in CMR is contingent upon the perceptual fusion evoked by sounds that come on synchronously or slightly asynchronously. From these CMR results, as well as those of a prior RMR experiment (Turgeon, 1999), the rhythm was expected to be released from masking whenever the irregular maskers and flankers were fully temporally overlapping, despite the combined action of many segregating cues: different ears of presentation and large ΔF 's. Though uncorrelated AM within the brief overlapping portions of the masker and flanker bursts was expected to diminish their fusion, and hence RMR, it was not expected to abolish it.

A. Methods

1. Subjects

There were 18 listeners who were naive to the purpose of the experiment. All listeners had normal hearing for the 250–8000 Hz frequency range, as assessed through a short air-conductance audiometric test.

2. Stimulus generation and presentation

All stimuli were synthesized and presented by a PC-compatible 486 computer, using MITSYN Version 8.1 signal processing software (Henke, 1990) and a 16-bit digital-to-analog converter. The rate of output was 20 000 samples per second. Signals were low-pass filtered at 5 kHz using a flat amplitude (Butterworth) response with a roll-off of 48 dB octave. Listeners sat in a sound-attenuating test chamber and listened to stimuli presented through Sony NR-V7 head-

phones. Stimuli were presented diotically, dichotically or monaurally, depending on condition. The rms level fluctuated slightly across the sequence due to random sampling of the noise components. The level of a 1-kHz pure tone equal in intensity to the mean rms of the noise bursts of the masked-rhythm sequence and of the flankers (measured as a pair) was calibrated at 60 dB SPL, using a General Radio Company Type 1565-B (“B” weighting, slow). The experiment was run on-line with the help of a MAPLE Version 2.0 program (Achim *et al.*, 1995) using ASYST Version 4.00 software.

3. Structure of sequences

Listeners were asked to discriminate two rhythms, presented as a sequence of noise bursts. These were made more difficult or impossible to discriminate by the insertion of maskers placed randomly in the time intervals between them. These were identical in all respects to the rhythmic components. Both of the rhythms were formed of the same set of long and short time intervals, but in a different arrangement. One cycle of each rhythm is shown in the left portion of panels (a) and (b) of Fig. 1. The long intervals were twice the duration of the shorter ones (short=384 ms; long=768 ms). Whereas Rhythm 1 repeated the sequence of intervals, short, long, short, and long, three and one-half times, Rhythm 2 repeated the sequence short, long, long, and short, three and one-half times. Figure 1 shows that there were two random maskers in the short interval, and four, in the long one. The temporal positions of the maskers were random from cycle to cycle. Except in the case of the no-flanker controls, these maskers were accompanied by noise bursts (“flankers”) situated in other frequency regions (see the white bars above and below the central sequence in the right portion of Fig. 1). The rhythm started at a variable time after the start of the irregular masking and flanking noise bursts and ended at a variable time before the irregular maskers and flankers stopped, keeping the total duration of the sequence constant across trials. This ensured that correct rhythm identifications did not result from the use of attentionally driven strategies exploiting local cues (e.g., listen for the short interval at the beginning of the sequence).

4. Structure of individual bursts

All the noise bursts (forming the masked-rhythm sequence and flankers) were 48-ms long, including a 8-ms quarter-sine onset and a 8-ms reversed-quarter-sine offset. Each burst was obtained by multiplication of an independent 1-to-100 Hz, 48-ms-long, nominally flat noise sample by a pure tone. This procedure yielded a 200-Hz-wide nominally flat noise band centered at the frequency of the tone. Each independent noise sample was created by the summation of closely spaced sinusoids (1-Hz apart) in randomly selected phases. The rhythmic and masker bursts were centered at 1700 Hz. The flankers were two 200-Hz-wide noise bands 48 ms in duration, equally distant from the central masking band. The ΔF between the maskers and each of the two flankers was either 619 Hz or 1238 Hz. Hence, the maskers

and flankers were always in different critical bands, as measured in equal rectangular bandwidths (Glasberg and Moore, 1990).

The amplitude fluctuation within each masker, due to the randomness of noise, could both be either correlated with that of its corresponding flanker burst, or not. The maskers and flankers that had correlated envelopes were obtained by using the same noise sample. This sample was multiplied by sinusoids of different frequencies to obtain masker and flanker bursts with different center frequencies. The random intensity changes were correlated throughout the overlapping portion of asynchronous maskers and flankers. This was produced by the method used to synthesize them. This involved the starting of the noise samples for the masker and flanker bursts at the same time, while triggering the gain control for the intensity of the delayed burst, only after the asynchrony time of a given condition. Different noise samples were used to obtain uncorrelated maskers and flankers. The order of presentation of the noise samples within the regular and irregular sequences was randomly varied across trials. Furthermore, a masker burst could be delayed or advanced relative to its two temporally adjacent flanking bands (there was an equal likelihood of each for any masker and flanker bursts). The amount of overlap between the maskers and flankers varied from full to none; that is, the SOA was either 0, 12, 24, 36, or 48 ms. The masked rhythm and the flankers could either be presented together to both ears (diotic) or separately to the two ears (dichotic). While for the former, the no-flanker control was diotic, for the latter, it was monotic.

5. Procedure

The listeners had to judge which of the two rhythms was embedded in the sequence and how clearly it was heard on a 5-point scale. They were instructed to use the lowest clarity rating of “1” when guessing. The other values of the scale corresponded to the following degrees of perceived clarity of the identified rhythm: “2” stood for “very unclear,” “3” for “unclear,” “4” for “clear,” and “5” for “very clear.” Listeners were familiarized with the procedure and brought to a high level of performance on non-masked sequences. They were then trained on masked sequences. To yield a stable performance, there were two practice sessions that provided feedback about accuracy of rhythm identification. Feedback continued to be provided throughout the subsequent sessions. The order of presentation of the different conditions was randomized across trials, except for the diotic and dichotic ones, which alternated across sessions. For the dichotic sessions, the listeners were instructed to direct their attention to one ear, namely that of the masked rhythm.

6. Design

a. Independent variables. The center frequencies of the masking and flanking bands were either 619 or 1238 Hz apart. The maskers and flankers were either presented in both ears or spatially separated through dichotic presentation; their asynchrony was 0, 12, 24, 36, or 48 ms. The temporal envelope of each masker was either correlated or not, with that of its overlapping flanker bursts. This was thus a 2×2

$\times 5 \times 2$ within-subject design with eight replications per cell. A no-flanker condition was added to verify that the rhythm was masked in the absence of any flanker.

b. Dependent variables. The accuracy of rhythm identification as well as the perceived clarity of the identified rhythm served as a measure of the fusion¹ of the flankers and maskers. This was based on the assumption that fusion was not all-or-none, but that higher degrees of fusion would lead to greater ease in distinguishing the rhythmic bursts from the irregular ones. We used the sensitivity measure of d' (MacMillan and Creelman, 1991, p. 8) and estimated SOA thresholds from the fitting of psychometric functions to proportion-correct (PC) scores (Weibull, 1951). We also used a measure that weighted the accuracy by the rated clarity of the identified rhythm; this weighted-accuracy scale (WA) was more sensitive to the effect of weak cues than were “pure objective measures of accuracy.” For instance, while the PC and d' scores did not show any significant difference for rapid envelope correlations, the WA scores revealed some significant ones for nearly synchronous stimuli. Because it captured weak effects well, WA was used to evaluate the relative weight of cues in favoring fusion. On the other hand, the fitting of Weibull functions to PC scores were more suited to evaluate the temporal limits for sound-event segregation. Lastly, the detection measure, d' , provided a conservative index of the most critical properties in cross-spectral fusion, since those were the ones most likely to affect the discriminability of the rhythm (signal). Individual response biases were also estimated (MacMillan and Creelman, 1991, p. 32).

B. Results and discussion

1. Measure of sensitivity to the target rhythm

Rhythm detectability and response bias (d' and c) were computed according to standard detection theory procedures.⁵ The WA scores were obtained by multiplying the absolute clarity rating of the listener on a 5-point scale by +1 when the rhythm was correctly identified, and by -1, when it was not. This yielded a scale ranging from -5 to +5. Because there were two units separating -1 from +1 (versus 1 unit between all other adjacent values of the scale), 0.5 was subtracted from the original clarity ratings to yield an equal-interval weighted-accuracy scale ranging from a -4.5 (incorrect “very clear”) to +4.5 (correct “very clear”). This equal-interval scale was required for the analysis of variance (ANOVA) of the WA scores, in the present case, a 4-way within-subject ANOVA. Note that none of the statistical assumptions to perform that ANOVA was violated, including that of the normal distribution of the WA scores.

Each individual subject's SOA threshold was determined from the best-fitting “Weibull” function (Weibull, 1951). Figure 2 shows Weibull psychometric functions for subject CB for the diotic conditions with a 619-Hz ΔF [continuous curve in panel (a)] and a 1238-Hz ΔF [dashed curve in panel (a)] and for the dichotic conditions with a 619-Hz ΔF [continuous curve in panel (b)] and a 1238-Hz ΔF [dashed curve in panel (b)]. Each of these functions minimizes the mean square estimate of error for the proportion of correct (PC) rhythm identifications as a function of SOA. For each con-

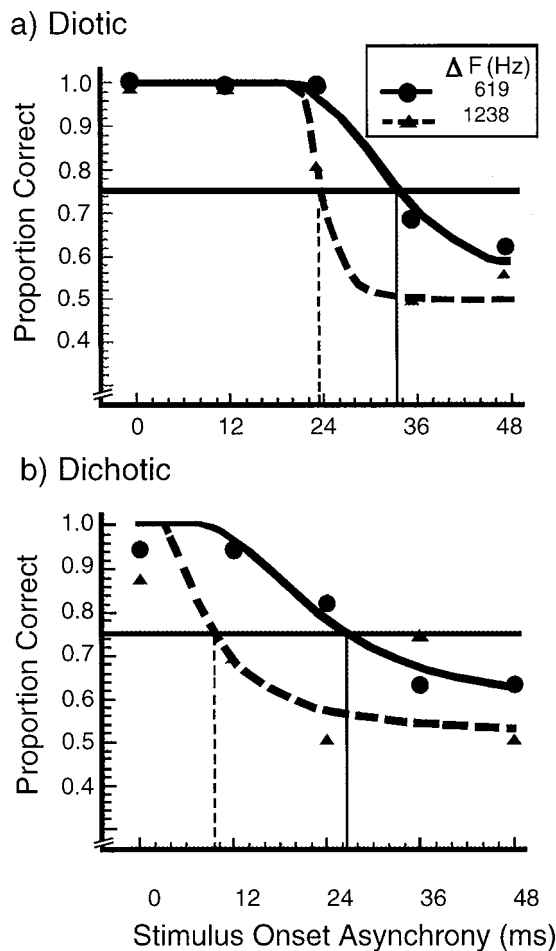


FIG. 2. Onset asynchrony (SOA) psychometric functions for subject CB for $\Delta F=619$ Hz (solid lines) and for $\Delta F=1238$ Hz (dashed lines), for diotic (a) and dichotic (b) conditions. SOA thresholds are taken as values yielding a proportion correct of 0.75.

dition, the mean of the within-subject Pearson-coefficient correlation (r) between the fitted function and the data points was at least 0.92. The threshold estimates were thus based on reasonably good fits of the PC data. Because there was very little difference between the PC scores obtained for the maskers and flankers with a correlated and uncorrelated envelope (mean PC of 0.72 and 0.69, respectively), the Weibull function was estimated from the PC scores collapsed across the two levels of envelope correlation.

The mean PC score obtained for the presentation of the masked rhythm alone (i.e., the no-flanker control) was 0.518, and the standard error (SE) for the 18 subjects was 0.015. The mean d' was 0.132, with a SE of 0.129 [see Fig. 3(b); the size of the SE corresponds to that of the cross symbol]. This performance was close to chance levels; hence, in the absence of flankers, the rhythm was perceptually masked. Given the continuous feedback about rhythm-identification accuracy, these results demonstrate that no attentionally driven strategies were able to overcome masking.

There was no evidence for individual bias towards either of the two rhythms ($0.5 < c < 0.5$), except for one listener who was biased towards Rhythm 2 ($c < -1$). Given that there was no consistent response bias for 17 out of 18 listen-

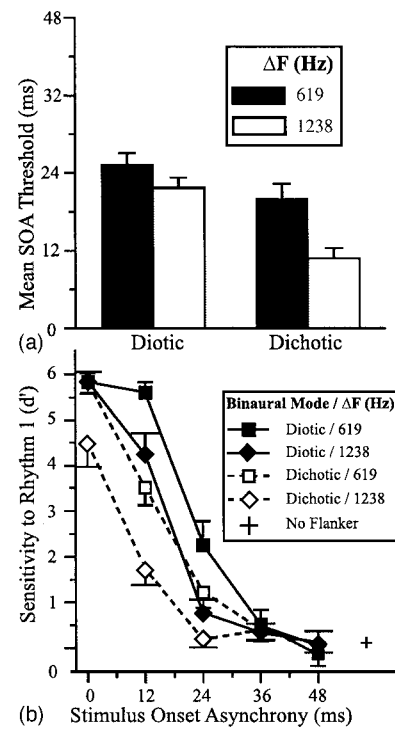


FIG. 3. (a) Mean onset asynchrony (SOA) thresholds for rhythm discrimination as a function of ΔF and binaural mode (ear of presentation). (b) Rhythm sensitivity as a function of ΔF and binaural mode. Higher thresholds and higher d' values represent higher degrees of fusion of the maskers and flankers. Data have been collapsed across the two levels of envelope correlation. Standard errors (SE) are shown; for the no-flanker control, its size corresponds to that of the cross symbol.

ers, it appears that the power of the statistical comparisons was not diminished by response bias.

2. Temporal resolution for event perception and rhythm discriminability

The threshold estimates of listener CB (see Fig. 2) were representative of those found for the 18 listeners. Panel (a) of Fig. 3 shows that the largest mean SOA threshold of 25.3 ms (SE=1.8 ms) was obtained for the diotic condition with the smaller 619-Hz ΔF . This was followed by the diotic condition with the larger ΔF of 1238 Hz (mean=21.8 ms, SE=1.5 ms), the dichotic condition with the 619-Hz ΔF (mean=20.1 ms, SE of 2.3 ms) and the dichotic condition with the 1238-Hz ΔF (mean=10.8 ms, SE=1.6 ms). Since a larger SOA threshold indicates that it was easier for the flanker to capture the masker into a common perceptual unit, it is concluded that cross-spectral fusion diminished with frequency separation as well as with the difference in the lateralization induced by dichotic presentation. The largest mean SOA threshold of 25.3 ms suggests that within the range of conditions of this experiment; an asynchrony of 25 ms triggers the perception of temporally contiguous sounds as separate events. However, smaller asynchronies can abolish their fusion in the presence of other cues, such as frequency separation and/or dichotic presentation. This is compatible with the near chance level of performance found for SOAs of at least 24 ms [see Figs. 3(b) and 4].

Figure 4 shows the WA scores as a function of SOA with ΔF [panel (a)], and envelope correlation and dichotic pre-

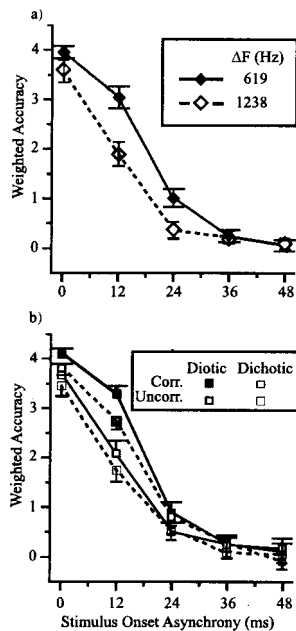


FIG. 4. Mean weighted accuracy (WA) as a function of SOA for maskers and flankers. (a) The parameter is ΔF , and the data have been collapsed across the two levels of envelope correlation and presentation (diotic/dichotic). (b) The parameters are presentation mode and envelope correlation, and the data have been collapsed across the two ΔF 's. Higher WA scores represent higher degrees of fusion of the maskers and flankers. The error bars represent ± 1 SE.

sensation [panel (b)] as parameters. As can be seen, the WA scores exhibit the same general trends as the rhythm-discrimination measure (d') shown in panel (b) of Fig. 3. However, while rapid correlated amplitude changes weakly increased fusion for periods of 36 and 48 ms of overlap [see solid versus dotted line at SOAs of 0 and 12 ms in Fig. 4(b)], they did not make the rhythm more discriminable than did their uncorrelated counterparts, as estimated by both PC and d' scores. Accordingly, the results in Fig. 3 have been collapsed across the two within-burst envelope correlations.

3. Relative contribution of cues to fusion: Description of the trends for WA scores

a. Interaction effects. Figure 4 shows that the weak effect of correlation of rapid intensity changes within individual bursts depended on SOA (only present for synchronous or nearly synchronous stimuli), on a large ΔF and on dichotic presentation. This is consistent with the four-way significant interaction, at the 5% level [$F(4,68) = 2.78$, $p = 0.03$]. Similarly, that the effect of ΔF depended on both the SOA value and dichotic presentation [see Fig. 3(b)] is reflected by the significant three-way interaction between these factors [$F(4,68) = 3.85$, $p = 0.007$]. Frequency separation interacted with both SOA [$F(4,68) = 18.07$, $p < 10^{-5}$] and the mode of presentation [$F(1,17) = 25.36$, $p = 0.0002$]. There was also a two-way interaction between SOA and the mode of presentation [$F(4,68) = 12.08$, $p < 10^{-5}$]. This indicates that the difference in WA between the diotic and dichotic stimuli depended on the value of SOA: though there was a difference in WA at 0 ms; it was larger at 12-ms SOA and was basically absent at 36-ms and 48-ms SOAs; this held at both ΔF 's [see Panels (a) and (b)

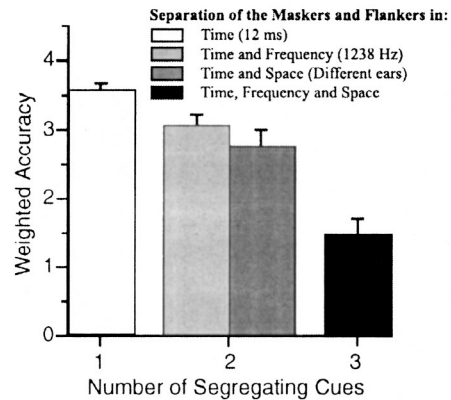


FIG. 5. Synergetic action of cues favoring segregation: a small SOA of 12 ms, dichotic presentation, and a large ΔF of 1238 Hz. The degree of fusion of the maskers and flankers separated by a 12-ms SOA decreases as the number of cues favoring segregation increases from one (white bar at left), to two (middle cluster of two bars) to three (black bar at right). For each listener, the mean WA scores were obtained from data collapsed across the two levels of envelope correlation.

of Fig. 4]. The fact that neither ΔF , nor dichotic presentation had an effect at SOAs of 36 and 48 ms is probably due to their near chance-level performances; that is, these asynchronies by themselves appear to abolish fusion. The results suggest that the effect of ΔF depended on the presence of other cues diminishing fusion: an asynchrony and/or dichotic presentation. Similarly, dichotic presentation, was not sufficient, unaided by asynchrony and/or a large ΔF , to abolish the fusion of simultaneous or nearly simultaneous sounds.

b. Main effects. Figure 4 shows that WA decreased with SOA [$F(4,68) = 381.22$, $p < 10^{-5}$]. This held for the diotic [$p < 10^{-5}$] and the dichotic [$p < 10^{-5}$] presentation of the masked rhythm and flankers. This suggests that SOA strongly affected fusion. Overall, WA also decreased with ΔF [$F(1,17) = 62.45$, $p < 10^{-5}$]; this held for the diotic [$p = 0.004$] and dichotic stimuli [$p < 10^{-5}$]. This figure also shows that cues reinforced each other in diminishing fusion.

Figure 5 provides another way to look at the interaction between cues which appear to weakly diminish fusion: a small temporal separation of 12 ms, a large frequency separation of 1238 Hz and spatial separation through dichotic presentation. In the absence of frequency and/or spatial separation, a 12-ms asynchrony only weakly diminished fusion (white bar); this is shown by its mean WA of 3.5, indicating that the correctly identified rhythm was rated on average as “clear.” Adding one of these cues (second set of bars), and the two of them (black bar) progressively diminished fusion more, to the point of almost abolishing it. The black bar shows a mean WA around 1.5, that is, when three cues acted together, the correctly identified rhythm was rated on average as “very unclear.” This suggests that a group of cues can have a synergetic action in diminishing fusion, though each by itself diminished fusion only weakly.

III. EXPERIMENT 2. TEMPORAL LIMITS AND RELATIVE CONTRIBUTION OF CUES TO THE CROSS-SPECTRAL FUSION OF NOISE BURSTS PRESENTED IN FREE FIELD

Experiment 2 was designed to generalize the results of experiment 1 to free field presentation. Using the RMR para-

digm, it explored the effects of temporal asynchrony, rapid envelope correlation and frequency separation (ΔF) on fusion. In addition, it manipulated the angular separation of their sources ($\Delta\theta$) and estimated temporal thresholds for event perception in a semi-circular speaker array.

In experiment 1, almost perfect rhythm identifications were obtained for synchronous masker and flanker bursts that were dichotically presented over headphones, had uncorrelated envelopes and were widely separated in frequency. Based on those results, it was expected that temporal coincidence would fuse the masker and flanker bursts, independently of values of the other cues. Experiment 1 also led us to expect that a SOA of 10–25 ms would lead to their perception as separate events, causing the rhythm to remain perceptually camouflaged. Compared to the slow intensity changes induced by temporal synchrony, the rapid ones within individual bursts were expected to have a negligible effect on fusion. It was also expected that ΔF and $\Delta\theta$ would weakly, but consistently, interfere with the perception of the rhythm; when present together, they should reinforce each other in diminishing fusion.

A. Methods

There were 18 normal-hearing listeners; 7 of them had also participated in experiment 1. The synthesis and presentation of the stimuli, as well as the procedure were the same as for experiment 1, except for a few points exposed below. The stimuli were presented from an array of 13 loudspeakers, situated in the sound-attenuated chamber of Dr. R. Zatorre, at the Montreal Neurological Institute (see Fig. 6). Listeners sat one meter away from each loudspeaker. The axis of the diameter of the semicircle, passing through the two end speakers (i.e., from 0 to 180 deg) passed through the axis of the two ears.

The mean rms level for the rhythmic and masking bursts and for the two flankers (measured as a pair) was calibrated to be equal to that of a 1-kHz tone presented over the central speaker and measured as 60 dB SPL at the central position of the listener's head. Due to the constraints of the available space and to keep the listeners' heads immobile, it was neither possible for them to record their responses directly into the computer after each trial, nor to read the computer screen for feedback and initiate new trials. Instead, the experimenter sat three meters away from the listener, behind the speaker array and close to the computer screen. From that position, she entered the listener's verbal response after each trial, read out the computer's feedback about whether or not the rhythm was correctly identified, and initiated each new trial. The listeners could neither see the speakers nor the experimenter during testing. At the beginning of each trial, a 1-kHz warning tone was played through the speaker of the masked rhythm, so that listeners could pay attention to its location. The listeners' heads remained fixed facing the central speaker, even when their attention was directed to other speakers.

Experiment 2 presented a new set of noise samples over loudspeakers, rather than over headphones as in experiment 1. Furthermore, while in experiment 1, the masker pulses could either precede or follow the corresponding flanker

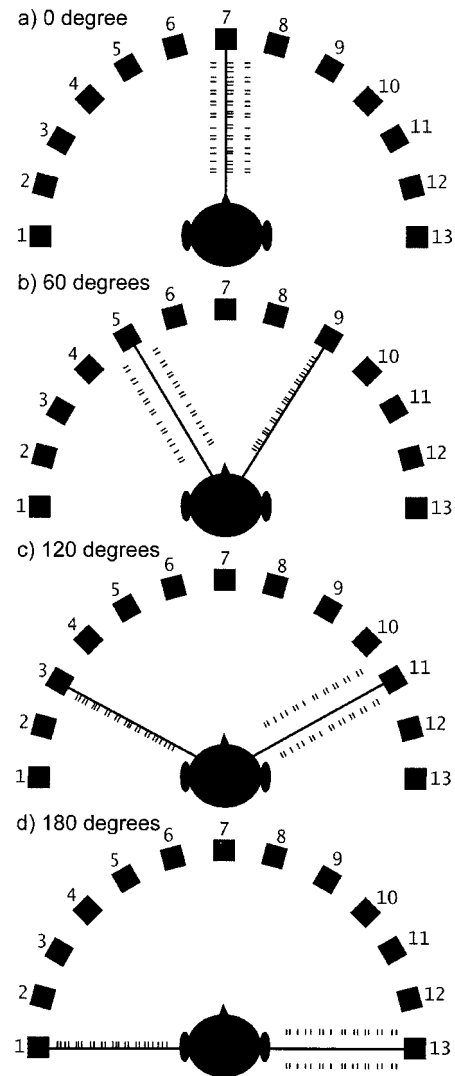


FIG. 6. Semi-circular array of 13 speakers used for the free-field presentation of the stimuli. The masked rhythm (illustrated by the single row of pulses) and the flankers (represented by the double row of pulses) could be presented in the central speaker of the array (a) or at various angular separations ($\Delta\theta$'s), namely, 60 deg (b), 120 deg (c), or 180 deg (d). For each $\Delta\theta$, the speakers of the masked speaker and flankers were symmetrically placed relative to the central speaker.

pulses, in experiment 2, the maskers always preceded the asynchronous flankers. Figure 6 shows that the masked rhythm and flankers could either be both presented in the central speaker [panel (a)] or at various angular separations ($\Delta\theta$) from it: 60, 120, and 180 deg [panels (b), (c), and (d)]. For each $\Delta\theta$, the masked rhythm and flankers came from speakers that were symmetrically placed relative to the central axis. The choice of which signal to present on each side of the array was counterbalanced across trials. There was a no-flanker control for each of the four $\Delta\theta$'s, in which the masked-rhythm sequence alone was either presented in the central speaker or at 30, 60, or 90 deg to the left or to the right of it. The different conditions were randomly presented across trials.

B. Results and discussion

1. Measures of sensitivity to the target rhythm

The degree of fusion of the maskers and flankers was assessed in the same way as for experiment 1. Because the

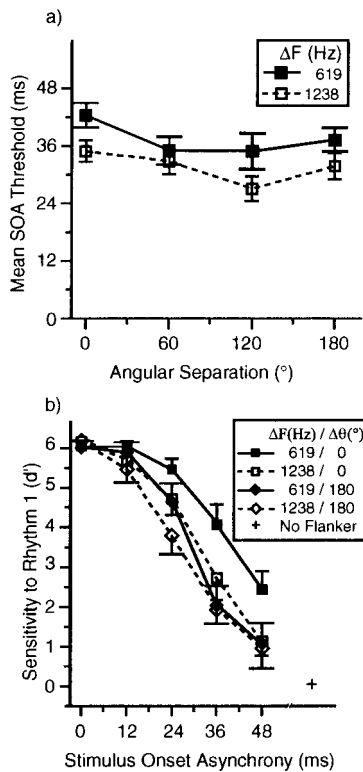


FIG. 7. (a) Mean onset asynchrony (SOA) thresholds for rhythm discrimination as a function of ΔF and $\Delta\theta$. (b) Rhythm sensitivity as a function of ΔF and $\Delta\theta$. Higher thresholds and higher d' values represent higher degrees of fusion of the maskers and flankers. Data have been collapsed across the two levels of envelope correlation. Standard errors (SE) are shown; for the no-flanker control, its size corresponds to that of the cross symbol.

mean PC scores, computed across listeners was 0.86 for correlated and 0.82 for uncorrelated AM (i.e., they differed by only 0.04), the Weibull functions were estimated from the PC scores collapsed across the two levels of envelope correlation. For each $\Delta\theta$ -by- ΔF condition, the mean, across listeners, of the within-subject Pearson-coefficient correlations (r) between the fitted function and the data points was at least 0.9.

2. No-flanker controls, performance range and response bias in rhythm-detection accuracy

The no-flanker control yielded a mean PC of 0.520 (SE of 0.018) and a mean d' of 0.059 [SE of 0.089 shown by the size of the cross symbol in Fig. 7(b)]. This very near chance-level performance verified that the rhythm was masked in the absence of flankers. On the other hand, synchrony fused spatially and spectrally distributed noise bursts. For each 0-ms SOA condition, the mean d' was larger than 4.65.

There was no systematic response bias for 17 listeners (i.e., $-0.2 > c < 0.2$). One listener had a slight bias towards Rhythm 1 for the conditions with and without flankers, the c values being -0.45 and -0.38 , respectively. Given that 95% of the listeners had very small c values, the power of the statistical comparisons was probably not diminished by response bias.

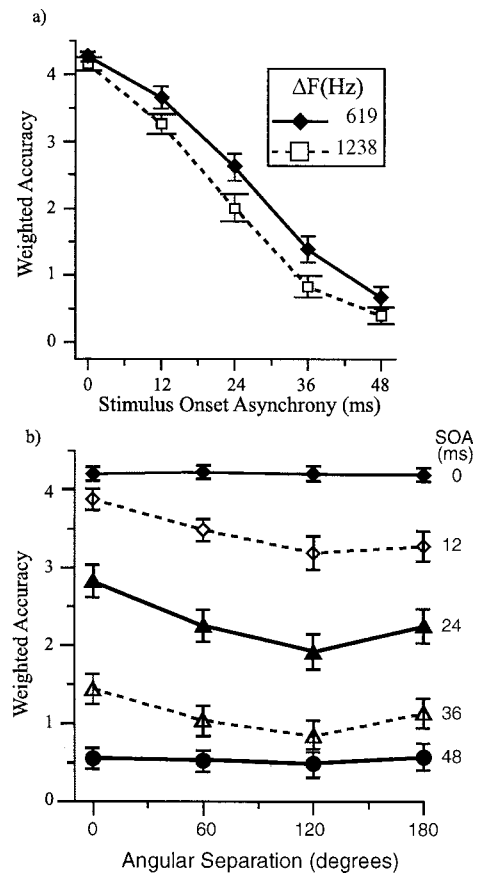


FIG. 8. Relative contribution of cues to the fusion of brief noise bursts separated in frequency and in space. Higher WA scores represent higher degrees of fusion. (a) WA as a function of SOA for the two ΔF 's collapsed across the two levels of envelope correlation and the four $\Delta\theta$'s. (b) WA as a function of $\Delta\theta$, with SOA as the parameter; data are collapsed across the two levels of envelope correlation and the two ΔF 's. The error bars represent ± 1 SE.

3. Temporal resolution for event perception and rhythm discriminability

The SOA-threshold estimates graphed in Fig. 7(a) show that a SOA between 25 and 45 ms (i.e., from one SE below the lowest mean threshold to one above the highest one) abolished fusion. The asynchrony required for the temporal resolution of brief noise bursts depended on how far apart they were in frequency [filled versus empty squares in Fig. 7(a)]. This was also the case in experiment 1 over headphones [black versus white bars in Fig. 3(a)]. This first experiment also showed higher thresholds for sounds presented in both ears versus different ears [left versus right bars in Fig. 3(a)]. This is consistent with the effect of spatial separation ($\Delta\theta$) on thresholds in this second experiment [Fig. 7(a)]. However, post-hoc pairwise comparisons showed that the effect of $\Delta\theta$ in diminishing fusion (i.e., lowering thresholds) was all-or-none: whether or not sounds came from the same speaker mattered, but how far apart the speakers were in space did not.

4. Relative contribution of cues: Description of the main trends for WA scores

Figure 8 shows a monotonic decrease in WA scores with increases in SOA [$F(4,68) = 324.46$, $p < 10^{-5}$]. On the

other hand, frequency separation (ΔF), spatial separation ($\Delta\theta$), and envelope correlation affected only weakly fusion. Their combined effect was not sufficient to overcome the powerful effect of synchrony on fusion. The lowest mean WA among the 0-ms SOA conditions was 4. It was obtained for both the 120 and 180 deg $\Delta\theta$'s with a 1238-Hz ΔF and uncorrelated envelopes; this value falls between 3.5 and 4.5 which correspond, respectively, to a "clear" and "very clear" rating of the correct rhythm. The effect of ΔF , $\Delta\theta$, and uncorrelated envelopes was clearest for the cases of partial fusion that were observed when the maskers and the flankers were partly overlapping (see Fig. 8). When the fusion of the maskers and flankers was complete, as at 0-ms SOA, or absent, as at 48-ms SOA, the contribution of a much weaker cue might be too small for its effect on fusion to be observable. This interpretation is consistent with the two-way interaction SOA-by- ΔF [$F(4,68)=4.18$, $p=0.005$] and SOA-by- $\Delta\theta$ [$F(12,204)=5.23$, $p<10^{-5}$] showing that the effects of both ΔF and $\Delta\theta$ depended on the value of SOA.

5. Unexpected effect of a sequential cue on fusion

Overall, the thresholds found when noise-burst stimuli were presented in a free field were higher than those found when they were presented over headphones [compare Figs. 3(a) and 7(a)]. A higher threshold means that fusion still took place at larger SOAs, fusion improving performance in the RMR task. The higher performance of the free-field presentation relative to that over headphones was also reflected by higher absolute mean WA scores across all conditions. This resulted in ceiling performance at 0-ms SOA and in near chance-level performance at 36-ms and 48-ms SOAs. The authors suspected that the common magnitude and direction of SOA in the global sequences of experiment 2—the two flanker bands were always delayed from the temporally adjacent masker by a given SOA—contributed to fusion over and above the local magnitude of SOA. To test whether such a sequential cue favored fusion, a post hoc analysis compared the performance at 48-ms SOA against that for the no-flanker control. Since there was no overlap between the maskers and flankers at that largest SOA, a higher degree of fusion could only be due to sequential cues. For each $\Delta\theta$, the 48-ms SOA condition yielded a higher WA than that obtained for the no-flanker control: the mean WA was almost equal across $\Delta\theta$'s, varying from 0.50 to 0.57; for the no-flanker control, it was only 0.03. The mean PC of 0.64 and d' of 0.74, obtained at 48-ms SOA were also larger than the PC of 0.51 and d' of 0.06 obtained for the no-flanker control. This can be contrasted with the near chance level of performance observed for the bi-directional 48-ms SOA condition of experiment 1, namely, mean PC of 0.54 and d' of 0.25. Post hoc comparisons between the 48-ms SOA and the no-flanker conditions were highly significant [$p<10^{-5}$] for WA, PC, and d' . This suggests that a constant direction of SOA affected the scores. It is also possible that an unforeseen difference between the free-field and binaural contexts affected performance.

C. Discussion

1. Relative contribution of temporal, spatial, and spectral separation to the segregation of temporally contiguous sound events

a. Strong effect of cross-spectral correlation of slow-varying intensity changes on fusion. The simultaneity of onsets and offsets across the masker and flanker bursts resulted in a slow pattern of correlated intensity changes across frequency regions. Given that such a pattern was nonperiodic (i.e., resulting from irregularly-spaced sounds), it did not have a frequency of modulation per se. However, one can say that overall it was slow, in that it resulted from sequences in which the temporal density of the irregular masker and flanker sounds was low, namely one per 192-ms interval. This is equivalent to 5.2 sounds/s and as such is comparable to a frequency of amplitude modulation of 5 Hz. When temporal synchrony, favoring fusion, competes with frequency separation and spatial separation (through presentation in different ears or loudspeakers), favoring segregation (i.e., diminishing fusion), synchrony played the determinant role, producing strong fusion (see Figs. 3, 4, 7, and 8). This powerful effect of synchrony is consistent with past results in the literature (see Darwin and Carlyon, 1995 for a review).

b. Very weak effect of cross-spectral correlation of fast-varying intensity changes on fusion. The slow correlation in intensity changes induced by temporal synchrony can be contrasted with that obtained through within-bursts envelope correlation. The latter is much more rapid and takes place over much shorter periods varying from 12 ms (36-ms SOA) to 48 ms (0-ms SOA) of temporal overlap. Such brief overlaps follow from the use of sounds of a constant 48-ms duration. The differences in WA resulting from envelope correlation were only observed at 0-ms and 12-ms SOA [solid versus dashed lines in Fig. 4(b)]. These results suggest that slow, but not fast intensity changes affect cross-spectral fusion. However, the fastest modulation in the envelope spectrum being nominally 100 Hz (i.e., the width of the common noise modulator), there might have been too few samples of the common amplitude envelope for the auditory system to reliably detect it. Future experiments should use noise bursts of a longer duration to determine whether there is a minimum number of cycles of the correlated waveform necessary to induce fusion. However, to truly compare the effect of the fast-varying amplitude changes of a local event to the slower ones of the global sequence on fusion, SOA should be manipulated independently from the duration of the sounds.⁶

c. Weak effect of large frequency separations on fusion. The weak, but consistent effect of frequency separation (ΔF) in this study replicates that found in a prior RMR experiment, with very similar stimuli (Turgeon, 1999). This experiment presented diotically and dichotically 200-Hz-wide noise bursts, the flanker bands being either 400, 550, 700, or 850 Hz remote in frequency from the masker bands centered at 1500 Hz. Furthermore, the role of ΔF in RMR is compatible with its effect on comodulation masking release or CMR (Hall *et al.*, 1984): though it did not abolish fusion, it reduced the degree of fusion of temporally overlapping, but asynchronous sounds (see Figs. 3 and 7). This effect has some ecological validity since causally related concurrent

sounds are more likely to cluster in frequency than causally unrelated ones. For instance, in many species, the sounds produced by a male and a female tend to be more distributed in frequency than those produced by a single female (the sounds produced by a female being typically in higher frequency registers than those of a male). However, being rich sounds, they are likely to partly overlap in frequency; hence the auditory system would still have to somehow separate the spectrally overlapped portions.

d. Weak effect of large spatial separations on fusion: Different ears versus different speakers. The weak but consistent effect of spatial separation in diminishing fusion suggests that it is used by the mammalian brain for sound-source determination. This is consistent with its role in promoting the identification of non-speech auditory patterns (Kidd *et al.*, 1998) and in the localization of concurrent sounds, as shown by studies on the concurrent minimum audible angle in a free field (Perrott, 1984) and in simulated space (Divenyi and Oliver, 1989). Taken together, these results suggest that spatial separation influences sound segregation (“how many”), identification (“what”), and localization (“where”), though it is not sufficient to segregate brief, concurrent, frequency-separated sounds (i.e., abolish their fusion). Figure 8 shows that unlike an asynchrony of 48 ms, which yielded a near-chance level of performance at each frequency and spatial separations (i.e., WA near 0), large spatial separations of 120 and 180 deg did not prevent the rhythm from being partly released from masking in the absence of an asynchrony of at least 36 ms. However, as Yost *et al.* (1996) have proposed, the separation of sources might play a more important role when more than two concurrent sounds are present. Further research should compare the contribution of spatial separation under conditions of varying number of concurrent sounds.

Figure 8 shows that the clearest effect of $\Delta\theta$ on the segregation of noise bursts is the contrast between sounds coming from the same speaker (a $\Delta\theta$ of 0 deg) or from different ones ($\Delta\theta$'s larger than 0 deg). These results suggest that the magnitude of the angular separation of sound sources is irrelevant for the segregation of sounds close together in time. A comparison between Figs. 3(a) and 7(a) suggests another important conclusion: the spatial disparity provided by dichotic presentation has more impact on the temporal resolution of brief concurrent sounds, than that provided by the spatial separation of their sound sources in a free field. It might be that dichotic separation is more efficient for sound segregation because it is an extreme case of interaural differences for sounds happening simultaneously, the stimulation of one sound being delivered to one ear only, while that of the other sound(s) is delivered to the other ear only. The free-field testing is more akin to real-world situations in which each of many individual sounds stimulates both ears,⁷ though at slightly different times and intensities, allowing for the computation of the location of each source. We suggest that when drawing conclusions about the contribution of spatial disparities, one should not consider dichotic presentation as reflecting ecologically valid differences in the location of sound sources. Even when two sound sources are close to different ears, a sound coming from one of them usually

stimulates the two ears, albeit with larger binaural differences in intensities and time of arrival than if sources were closer to the midline axis. For this reason, the separation of sound sources in a free field is considered as more representative of the true contribution of spatial separation to sound-source segregation. This contribution is weak when two sources are simultaneously active. Further experimentation should determine whether these conclusions apply to sounds of a longer duration, as well as to temporally contiguous but non overlapping sounds.

2. Binaural fusion is not spectrally limited for the purpose of sound-source determination

Taken together, the results from dichotic RMR and CMR provide evidence for a process that performs a cross-spectral analysis of the low-rate amplitude changes. It is suggested that this analysis generalizes to the way in which acoustic information that is spread out over the spectrum or over space will contribute to the perception of either a single sound or more than one sound. Contrary to this suggestion, some experiments have concluded that binaural fusion has clear spectral limits (Perrott and Barry, 1969; van den Brink *et al.*, 1976). Perrott and Barry (1969) have shown that the fusion of concurrent pure tones presented to the different ears is contingent upon them having less than a critical difference in frequency, which is proportional to the frequency of the tones themselves (approximately 4% of the latter). Other experiments on the concurrent minimum audible angle (CMAA) in simulated space (Divenyi and Oliver, 1989) and in a free field (Perrott, 1984) have demonstrated that the auditory system has a very poor spatial resolution (as much as 60 deg) for spectrally overlapping concurrent sounds as well as for spectrally nonoverlapping sounds that are close in frequency. Assuming that poor spatial resolution is linked to poor perceptual segregation, this provides supportive evidence that the binaural segregation of spectrally overlapping sounds requires a wide spatial separation. This interpretation is further reinforced by the work of Scharf *et al.* (1976) which suggests that the segregation of two components in space is most likely to occur when their spectral patterns show little or no overlap. Given that fusion is the absence of segregation,¹ if sounds are not segregated in space, as in CMAA, they must be fused, at least partly. Therefore, together, the results on binaural fusion and on the CMAA provide evidence that the sound-source determination of concurrent sounds is somehow spectrally limited. How can these results be reconciled with the results obtained with the RMR and CMR paradigms, which together provide evidence for a cross-spectral binaural analysis underlying sound-source determination?

The present research suggests that sounds coming from different locations in space can be perceived as a single environmental event, without their being spectrally matched (spectral matching being typical of sounds arising from a common natural source). This is consistent with past observations which mention that spectrally remote sounds can evoke a single image, though it is typically not well localized and described as “diffuse” (Perrott and Barry, 1969; Thurlow and Elfner, 1959). In the present study, the observations

of the authors suggest that the fusion of the maskers and flankers which were perceived as being causally related, but were spectrally and spatially remote, produced the localization of the masker-flanker complexes to either a virtual source, or to the veridical source of the flankers. The different tasks used in the RMR paradigms, CMAA studies, and other studies of binaural fusion might have been looking at different types of fusion: (i) The fusion of spectrally overlapping components distributed in space through binaural cross-correlation localization mechanisms (Jeffress, 1972; Lindemann, 1986); and (ii) The fusion of spectrally nonoverlapping components through independent pre-attentive grouping processes.¹ When fusion is defined as the perception, as a single sound event, of many frequency components that might or might not be distributed in space, there does not seem to be any spectral limit for fusion. On the other hand, the perception of a single sound event at a definite location in the environment (“what is where”) appears to be spectrally limited (Divenyi and Oliver, 1989; Perrott, 1984; Scharf *et al.*, 1976).

3. Implications of the results for the psychophysical limits of event perception

The two RMR experiments suggest that the asynchrony needed for the segregation of brief sound events with abrupt onsets and offsets is about 20-to-40 ms; however, it can be lowered by the synergetic action of other simultaneous-grouping cues such as frequency and spatial separation, as well as sequential ones, such as a constant direction of asynchrony. The range of asynchrony thresholds is in general agreement with the literature on auditory grouping showing that an asynchrony of 30–40 ms is required for removing a partial from contributing to the overall timbre (Bregman and Pinker, 1978), to the lateralization (Hill and Darwin, 1993) or to the vowel identity (Darwin, 1981) of a complex sound. If timbre, vowel quality, and perceived lateralization are assumed to be properties of perceptually segregated sounds, one should expect this close correspondence. The 20-to-40 ms asynchrony in these phenomena is about an order of magnitude higher than the 2–3 ms required for the cross-spectral detection of an asynchrony (Green, 1973) and an order of magnitude lower than the 200–300 ms asynchrony preventing a partial from contributing to the pitch of a complex tone (Darwin and Ciocca, 1992). The fact that a just detectable asynchrony is not sufficient to segregate temporally overlapping sounds in different parts of the spectrum is compatible with the observation that listeners report only a single click while reliably detecting very small asynchronies (Green, 1973). At the other extreme, the discrepancy between the temporal limits for pitch and for event perception might indicate different underlying neural mechanisms. The temporal limits for event perception should be further investigated, especially as they relate to other spectro-temporal regularities known to influence auditory organization, both of local properties of a sequence (e.g., duration and rate of onset and offset of each sound) and global ones (e.g., tempo, distribution of silent intervals). Further experiments should also look at how time-varying intensity changes interact with spectral

regularities influencing the computation of pitch and the fusion of the tonal sounds forming harmonic and inharmonic complexes (Roberts and Brunstrom, 1998).

4. Implications of the results for the probabilistic approach to event perception

In complex world environments, many sources of acoustic evidence typically converge upon a common perceptual interpretation: sounds come from the same spatial location, start and stop at the same time and undergo common spectro-temporal changes. The present study created competition among alternative auditory organizations. Such ambiguous stimuli unveil the relative importance of grouping cues. In this study, despite the combined effect of frequency and spatial separation in favoring segregation, simultaneous sounds fused strongly enough to perceptually release the rhythm from masking. The possibility that temporal synchrony is “weighted” more strongly than frequency and spatial separation taken together has ecological validity. Temporal coincidence is a highly reliable and robust property of the components of biologically relevant sounds. Although it is likely for sounds coming from different sources to have some degree of temporal overlap, it is highly unlikely that they happen to start and stop at exactly the same time. On the other hand, frequency separation is not as reliable a cue since concurrent sounds coming from a common biological source typically occupy different frequency regions (Yost and Sheft, 1993); conversely those coming from different sources can overlap in frequency. Similarly, causally related sounds need not have a sharply focused location, because they go through and around some surfaces and are reflected by others. Because of these properties of the acoustic world, onset synchrony and deviations from it are more informative to a biological system than either the frequency or the spatial separation of acoustic components. This might explain why asynchrony contributed more to sound segregation than other cues did in the present study. Such a weighted contribution might, however, be dependent on the particular methodology of RMR studies as well as the parameters values used in these experiments. Experiments with other tasks and stimuli should look at the issue to determine to what extent these results are generalizable to the perceptual organization of complex sounds.

ACKNOWLEDGMENTS

This research was supported in part, by a grant from the National Sciences and Engineering Research Council of Canada (NSERC) to A. S. Bregman and in part by the FCAR program of the Province of Quebec. We are grateful for the technical assistance of Pierre A. Ahad of the McGill Auditory laboratory. We also want to thank Dr. Robert Zatorre for having allowed us to do the free-field testing in his laboratory at the Montreal Neurological Institute.

¹Glossary of terms: *Auditory grouping* involves the perceptual organization of sound components into coherent perceptual units. These units can be isolated sound events (e.g., a hand clap) or sequences of them (e.g., a melody), also known as streams. The term “grouping” typically encompasses both the *fusion* of *n* sound components (usually concurrent ones)

into a single event and their *segregation* as n distinct sounds (sequential or concurrent ones). *Pre-attentive auditory grouping* results from biologically implemented processes exploiting environmental regularities with adaptive value; this is independent of acquired knowledge, not under the control of attentional mechanisms and can be contrasted with attentionally driven grouping (Bregman, 1990). *Cross-spectral integration* is the summing up of time-varying intensity changes across frequency-selective channels; the information might or might not be weighted equally in different channels. *Cross-spectral fusion* happens when cross-spectral integration leads to the perception of multiple frequency components as a single sound event.

²We conceive of fusion and segregation as being the two extremes of a continuum: at one extreme, many sound components are fused, that is, they are perceived as one sound event; at the other extreme, they are segregated, that is, perceived as separate sound events. There are intermediate cases of *partial fusion* on this continuum in which many sound components can be perceived as many sound events or as a single one with global properties different from those of its parts. For such ambiguous percepts, what is heard depends on attentional factors and/or cognitive expectations, such as trying to hear out the individual notes of a chord versus the whole chord. Although segregation is assumed to be inversely correlated with fusion, for clarity purposes, the results have been described mainly, in terms of “degrees of fusion,” such as complete fusion, partial fusion, no fusion (i.e., segregation). This is not to imply that we take a position about whether or not fusion is the default for simultaneous sounds. In the statements of hypotheses and interpretation of results, we refer to both “fusion” and “segregation;” the use of each term being justified by how a cue is typically described in the literature (e.g., a segregation cue, in the case of onset asynchrony and a fusion cue, in the case of a common fundamental frequency).

³In RMR, a sequence is masked but its individual sounds are not. This type of masking can be distinguished from that due to peripheral signal-to-noise constraints (Zwicker, 1970). Such *energetic masking* takes place between temporally overlapping sounds that are either spectrally overlapping or close in frequency (e.g., when a noise burst perceptually masks a simultaneously present sinusoidal signal situated in the same frequency region, as in CMR). The *sequential masking* in RMR is more akin to *informational masking* (IM) in which the signal and maskers are perceptually segregated objects (Kidd *et al.*, 1994). Unlike IM, in which the release from masking can be due to spectral or temporal regularities among the subset of sequential target components that are not shared by the masking ones, in RMR, no sequential property distinguishes the components of the target rhythm from those of the masking sequence (i.e., apart from their regular vs irregular arrangement). Rather, the rhythm is released from masking by spectro-temporal regularities relating the concurrent maskers and flankers that cause them to fuse perceptually.

⁴This effect of frequency separation between the on-signal masking band and the flanking bands might be inflated by within-channel processes, such as “dip listening” (Schooveveldt and Moore, 1987).

⁵Sensitivity to Rhythm 1 (d') and response biases (c) were estimated for each subject from signal detection theory. In terms of Z (i.e., the inverse of the normal distribution function), d' is defined as $Z(H) - Z(F)$ and c , as $0.5 * [Z(H) + Z(F)]$; where “ H ” is the proportion of Hits and “ F ” is the proportion of False Alarms. In the present experiments, “ H ” was the proportion of correctly identified Rhythm 1 and “ F ,” that of falsely identified Rhythm 1 (i.e., Rhythm 2 was present). To avoid values of infinity in the computation of d' , proportions of 1 were converted into 0.999; this yields a d' value of 6.18. Hence, we used a d' value of 0 as the chance-level performance, and a d' of 6.18 as the perfect performance. When the frequency of response to Rhythm 1 across all trials is equal to that of Rhythm 2, the response bias statistic (c) equals zero (i.e., False Alarm and Miss rates are equal). A positive or a negative c indicates a higher tendency to respond “Rhythm 1” or “Rhythm 2,” respectively.

⁶The lack of discriminability of fast-varying amplitude correlations might be responsible for the lack of an RMR effect. Another concern is that the period of within-burst correlation decreased with asynchrony. Consequently, part of the SOA effect might reflect the reduction of fine-scale envelope correlation. This follows from the constant duration of the bursts, which allowed for a wider range in the selection of the silent intervals between irregular bursts. Despite the possible contribution of the correlated-envelope duration to the overall effect of SOA (half of the masker-flanker events being correlated), there was a clear effect of SOA on fusion. SOA had a strongly significant effect [$p < 10^{-5}$], both for the conditions of this experiment, and in a prior experiment in which the maskers and flankers were always uncorrelated (Turgeon, 1999).

⁷There are some rare exceptions to this generalization; for instance, the sound produced when there is an insect in one ear stimulates only the receptors in that ear.

- Achim, A., Bregman, A. S., and Ahad, P. A. (1995). “Manager of Auditory Perception and Linguistic Experiments (MAPLE).” Montreal QC., Canada: Auditory Perception Laboratory, Dept. of Psychology, McGill University.
- Bregman, A. S. (1978). “Auditory streaming: Competition among alternative organizations,” *Percept. Psychophys.* **23**, 391–398.
- Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA).
- Bregman, A. S. (1993). “Auditory scene analysis: Hearing in complex environments,” in *Thinking in Sounds: Cognitive Aspects of Human Audition*, edited by S. McAdams and E. Bigand (Oxford University Press, Oxford), pp. 10–36.
- Bregman, A. S., and Ahad, P. (1996). “Demonstrations of Auditory Scene Analysis: The perceptual organization of sound,” (Compact disk and booklet, pp. 41–42), Auditory Perception Laboratory, Psychology Dept., McGill University.
- Bregman, A. S., and Pinker, S. (1978). “Auditory streaming and the building of timbre,” *Can. J. Psychol.* **32**, 19–31.
- Bregman, A. S., Abramson, J., Doehring, P., and Darwin, C. J. (1985). “Spectral integration based on common amplitude modulation,” *Percept. Psychophys.* **37**, 483–493.
- Brochard, R., Drake, C., Botte, M.-C., and McAdams, S. (1999). “Perceptual organization of complex auditory sequences: Effect of number of simultaneous subsequences and frequency separation,” *J. Exp. Psychol., Hum. Percept. Perform.* **25**, 1742–1759.
- Dannenbring, G., and Bregman, A. S. (1978). “Streaming vs fusion of sinusoidal components of complex tones,” *Percept. Psychophys.* **24**, 369–376.
- Darwin, C. J. (1981). “Perceptual grouping of speech components differing in fundamental frequency and onset-time,” *Q. J. Exp. Psychol.* **33**, 185–207.
- Darwin, C. J., and Carlyon, R. (1995). “Auditory grouping,” in *The Handbook of Perception and Cognition, Volume 6, Hearing*, 2nd ed., edited by B. C. J. Moore (Academic, London), pp. 387–424.
- Darwin, C. J., and Ciocca, V. (1992). “Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component,” *J. Acoust. Soc. Am.* **91**, 3381–3390.
- Divenyi, P. L., and Oliver, S. K. (1989). “Resolution of steady-state sounds in simulated auditory space,” *J. Acoust. Soc. Am.* **85**, 2042–2052.
- Glasberg, B. R., and Moore, B. C. J. (1990). “Derivation of auditory filter shapes from notch-noise data,” *Hear. Res.* **47**, 103–138.
- Gordon, P. C. (1997). “Coherence masking protection in brief noise complexes: Effects of temporal patterns?” *J. Acoust. Soc. Am.* **102**, 2276–2283.
- Green, D. M. (1973). “Temporal acuity as a function of frequency,” *J. Acoust. Soc. Am.* **54**, 373–379.
- Green, D. M. (1988). *Profile Analysis: Auditory Intensity Discrimination* (Oxford University Press, New York).
- Grose, J. H., and Hall III, J. W. (1993). “Comodulation masking release: Is comodulation sufficient?” *J. Acoust. Soc. Am.* **93**, 2896–2902.
- Hall III, J. W., and Grose, J. H. (1991). “Some effects of auditory grouping factors on modulation detection interference (MDI),” *J. Acoust. Soc. Am.* **90**, 3028–3035.
- Hall III, 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.
- Handel, S. (1989). “Rhythm,” Chapter 5, in *Listening: An Introduction to the Perception of Auditory Events* (MIT Press, Cambridge, MA).
- Hartmann, W. M. (1988). “Pitch perception and the organization and integration of auditory entities,” in *Auditory Function: Neurobiological Bases of Hearing*, edited by G. W. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 623–645.
- Henke, W. L. (1990). *MITSYN Languages* (Belmont, MA).
- Hill, N. I., and Darwin, C. J. (1993). “Lateralization of a perturbed harmonic: Effects of onset asynchrony and mistuning,” *J. Acoust. Soc. Am.* **100**, 2352–2364.
- Hukin, R. W., and Darwin, C. J. (1995). “Effects of contralateral presentation and of interaural time differences in segregating a harmonic from a vowel,” *J. Acoust. Soc. Am.* **98**, 1380–1387.
- Jeffress, L. A. (1972). “Binaural signal detection: Vector theory,” in *Foundations of Modern Auditory Theory, Vol. II*, edited by J. V. Tobias (Academic, New York).

- Kidd, G., Mason, C., Rohdla, T. L., and Deliwala, P. S. (1998). "Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns," *J. Acoust. Soc. Am.* **104**, 422–431.
- Kidd, G., Mason, C., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (1994). "Reducing informational masking by sound segregation," *J. Acoust. Soc. Am.* **95**, 3475–3480.
- Lindemann, W. (1986). "Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals," *J. Acoust. Soc. Am.* **80**, 1608–1622.
- Macmillan, N. A., and Creelman, C. D. (1991). *Detection Theory: A User's Guide* (MIT Press, Cambridge, MA).
- McFadden, D. M. (1986). "Comodulation masking release: Effects of varying the level, duration, and time delay of the cue band," *J. Acoust. Soc. Am.* **80**, 1658.
- McFadden, D. M., and Wright, B. A. (1990). "Temporal decline of masking and comodulation detection differences," *J. Acoust. Soc. Am.* **88**, 711–724.
- Moore, B. C. J. (1989). *An Introduction to the Psychology of Hearing*, 3rd ed. (Academic, New York).
- Perrott, D. R. (1984). "Concurrent minimum audible angle: A re-examination of the concept of auditory spatial acuity," *J. Acoust. Soc. Am.* **75**, 1201–1206.
- Perrott, D. R., and Barry, S. H. (1969). "Binaural fusion," *J. Aud. Res.* **9**, 263–269.
- Roberts, B., and Brunstrom, J. M. (1998). "Perceptual segregation and pitch shifts of mistuned components in harmonic complexes and in regular in-harmonic complexes," *J. Acoust. Soc. Am.* **104**, 2326–2338.
- Scharf, B., Florentine, M., and Meiselman, C. H. (1976). "Critical band in auditory lateralization," *Sens. Processes* **1**, 109–126.
- Schooneveldt, G. P., and Moore, B. C. J. (1987). "Comodulation masking release (CMR): Effects of signal frequency, flanking band frequency, masker bandwidth, flanking band level, monotic versus dichotic presentation of the flanking band," *J. Acoust. Soc. Am.* **82**, 1944–1956.
- Sheft, S., and Yost, W. A. (1997). "Binaural modulation detection interference," *J. Acoust. Soc. Am.* **102**, 1791–1798.
- Steiger, H., and Bregman, A. S. (1982). "Competition among auditory streaming, dichotic fusion and diotic fusion," *Percept. Psychophys.* **32**, 153–162.
- Thurlow, W. R., and Elfner, L. F. (1959). "Pure-tone cross-ear localization effects," *J. Acoust. Soc. Am.* **31**, 1606–1608.
- Turgeon, M. (1999). Chapter 2 in: *Cross-spectral auditory grouping using the paradigm of rhythmic masking release: Doctoral dissertation*, McGill University, Montreal, Qc., Canada.
- Turgeon, M. (1994). *The influence of log-frequency parallel gliding upon perceptual fusion*. Master's thesis, McGill University, Montreal, Qc., Canada.
- Turgeon, M., and Bregman, A. S. (1997). "'Rhythmic Masking Release: A paradigm to investigate auditory grouping resulting from the integration of time-varying intensity levels across frequency and across ears,'" *J. Acoust. Soc. Am.* **102**, 3160.
- van den Brink, G., Sintnicolaas, K., and van Stam, W. S. (1976). "Dichotic pitch fusion," *J. Acoust. Soc. Am.* **59**, 1471–1476.
- Weibull, W. A. (1951). "A statistical distribution function of wide applicability," *J. Appl. Mech.* **18**, 292–297.
- Woods, W. S., and Colburn, H. S. (1992). "Test of a model of auditory object formation using intensity and interaural time difference discrimination," *J. Acoust. Soc. Am.* **91**, 2894–2902.
- Yost, W. A. (1991). "Auditory image perception and analysis: the basis for hearing," *Hear. Res.* **56**, 8–18.
- Yost, W. A., and Sheft, S. (1993). "Auditory perception," in *Human Psychophysics*, edited by W. A. Yost, A. N. Popper, and R. F. Fay (Springer-Verlag, New York), pp. 193–236.
- Yost, W. A., Dye, Jr., R. H., and Sheft, S. (1996). "A simulated 'Cocktail Party' with up to three sound sources," *Percept. Psychophys.* **58**, 1026–1036.
- Yost, W. A., Sheft, S., and Opie, J. (1989). "Modulation interference in detection and discrimination of amplitude modulation," *J. Acoust. Soc. Am.* **86**, 2138–2147.
- Zwicker, E. (1970). "Masking and psychological excitation as consequences of the ear's frequency analysis," in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (AW Sijthoff, The Netherlands), pp. 376–394.