Journal of Experimental Psychology:
Human Perception and Performance
2005, Vol. 31, No. 5, 939–953

Copyright 2005 by the American Psychological Association
0096-1523/05/$12.00 DOI: 10.1037/0096-1523.31.5.939

Rhythmic Masking Release: Effects of Asynchrony, Temporal Overlap, Harmonic Relations, and Source Separation on Cross-Spectral Grouping

Martine Turgeon
University of Birmingham

Albert S. Bregman
McGill University

Brian Roberts
University of Birmingham

The rhythm created by spacing a series of brief tones in a regular pattern can be disguised by interleaving identical distractors at irregular intervals. The disguised rhythm can be unmasked if the distractors are allocated to a separate stream from the rhythm by integration with temporally overlapping captors. Listeners identified which of 2 rhythms was presented, and the accuracy and rated clarity of their judgment was used to estimate the fusion of the distractors and captors. The extent of fusion depended primarily on onset asynchrony and degree of temporal overlap. Harmonic relations had some influence, but only an extreme difference in spatial location was effective (dichotic presentation). Both preattentive and attentionally driven processes governed performance.

Keywords: auditory grouping, stream segregation, sound-source separation, perceptual organization, event synchrony

General Background

The perceptual grouping of sounds depends on the detection of biologically relevant cues in the acoustic signal, such as those reflecting spectrotemporal regularities typical of causally related sounds. There is evidence that the auditory system groups perceptually sounds that share a common fundamental frequency (Demany & Semal, 1988; Hartmann, McAdams, & Smith, 1990; Moore, Glasberg, & Peters, 1986; Roberts & Brunstrom, 1998) and a common spatial location (Hukin & Darwin, 1995; Kidd, Mason, Rohtla, & Deliwala, 1998; Yost, 1997; Yost, Dye, & Sheft, 1996). Other research has shown that perceptual integration of acoustic elements is favored if they begin and end at the same time (Dannenbring & Bregman, 1978; Darwin & Sutherland, 1984; Roberts & Moore, 1991; Turgeon, Bregman, & Ahad, 2002). Conversely, the segregation of sounds is promoted by deviations from temporal coincidence, from simple harmonic ratios, and from a common source location. Though the contribution of many specific cues to auditory grouping has been established empirically (for reviews, see Bregman, 1990, 1993; Carlyon, 2004; Darwin, 1997; Darwin & Carlyon, 1995; Yost, 1991), their interaction is poorly understood, especially in a free field. It is important to study grouping in the context of many interacting cues, because no grouping cue acts in isolation in real-world situations. Rather, many cues interact—sometimes competing with each other, sometimes reinforcing each other—to provide the animal with a useful perceptual description of the environment.

The present study aimed to evaluate the relative contribution of spectral, spatial, and temporal auditory-grouping cues and their interactions. Toward that goal, the rhythmic masking release (RMR) paradigm was used (Bregman & Ahad, 1996; Turgeon & Bregman, 1997; Turgeon et al., 2002). The RMR study by Turgeon et al. (2002) used sequences of narrow-band noises, but the current study used sequences of complex tones, so that harmonic relations could be included as a factor. The following section illustrates how RMR is a useful paradigm for studying cross-spectral fusion in the context of competition between alternative auditory organizations.

Rationale of the RMR Paradigm

An illustrative example is useful to explain the rationale of the RMR paradigm as it applies to the present study. Figure 1a shows that a regular sequence of four-partial tones, comprising the succession of a short interval with a long one of twice its duration, is perceived as a sequence of rhythmic pairs of tones. When tones acoustically identical to the latter are interspersed irregularly

---

Martine Turgeon, Sensory Motor Neuroscience Laboratory, School of Psychology, University of Birmingham, Edgbaston, Birmingham, England; Albert S. Bregman, Auditory Perception Laboratory, Department of Psychology, McGill University, Montreal, Quebec, Canada; Brian Roberts, Auditory Perception Laboratory, School of Psychology, University of Birmingham.

Part of this research was presented in a PhD thesis submitted to the Department of Psychology of McGill University by Martine Turgeon under the supervision of Albert S. Bregman. Support was provided in part by a grant from the National Sciences and Engineering Research Council of Canada to Albert S. Bregman and in part from a Medical Research Council (United Kingdom) cooperative group grant on which Brian Roberts is a coinvestigator. We thank Pierre Ahad for his technical assistance and Robert Zatorre for allowing the free-field testing to take place in his laboratory at the Montreal Neurologic Institute. We also thank Michael Kubovy for his comments on an earlier version of this article.

Correspondence concerning this article should be addressed to Martine Turgeon, Sensory Motor Neuroscience Laboratory, School of Psychology, University of Birmingham, Edgbaston B15 2TT, Birmingham, England. E-mail: m.turgeon@bham.ac.uk

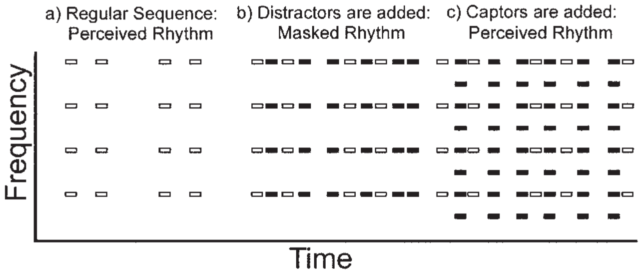

Figure 1. Rhythmic masking release with complex tones. a: A regular sequence of complex tones (open rectangles) is perceived as rhythmic when presented in isolation. b: Embedding an irregular sequence of identical tones as distractors (filled rectangles) within this regular sequence perceptually masks the rhythm. c: The rhythm is released from masking when harmonically related tones of different frequencies (captors) are added simultaneously to the distractors (additional filled rectangles not preceded by open rectangles). Note that this schematic is not intended to reflect accurately the time intervals between the rhythmic tones or the number of distractors inserted in these intervals.

among them, the rhythm is no longer heard (see Figure 1b). The rhythm is perceptually camouflaged, because no acoustic property distinguishes the regular subset of tones (open rectangles) from the irregular one (filled rectangles). We refer to the regular tones as *rhythmic targets* (Ts) and to the irregular tones as *distractors* (Ds). Although the Ds do not mask the individual Ts, they mask their rhythmic sequential organization (see Figure 1a). The whole sequence shown in Figure 1b is referred to as the *masked rhythm.*

Captor (C) tones can be added simultaneously with the irregular Ds (these are the filled rectangles that have different frequency components from the tones of the masked rhythm in Figure 1c). The addition of these C tones should release the rhythm from masking if the common onset and offset times cause them to fuse perceptually with the irregular Ds (this is illustrated by the vertical grouping of the filled rectangles in Figure 1c). The newly formed distractor–captor units (DCs) have changed properties, such as a new timbre and pitch. In our example, the DCs have eight partials and the rhythmic tones have only four; furthermore, the low pitch of the DCs is an octave below that of the rhythmic tones. The perceptual fusion of the Ds and Cs distinguishes the components of the irregularly spaced sounds (the filled rectangles) from those of the regularly spaced ones (the open rectangles). The accurate perception of the rhythm is thus contingent on the cross-spectral fusion of the irregular Ds and Cs. Therefore, measuring the listener's ability to identify the embedded rhythm can provide an estimate of the degree of perceptual fusion of the Ds and Cs. We manipulated the spatial, spectral, and temporal relations between the Ds and Cs to see how their fusion was affected by these factors using a two-alternative single-interval task in which one of two rhythms was embedded in the sequence.

This paradigm can be seen as involving competition between alternative perceptual organizations. Suppose that the masked rhythm sequence and the Cs of Figure 1 are presented in different speakers. Whereas the shared F0 (fundamental frequency giving rise to low pitch) and speaker location of the Ts and Ds should promote their sequential grouping, the temporal coincidence and harmonic relations between the Ds and Cs should promote their simultaneous grouping. If common location and pitch overcome the grouping effects of temporal coincidence and harmonic relations, the rhythm should remain perceptually masked. However, if temporal coincidence and harmonic relations (among spatially and spectrally distributed components) win the competition, the Ds and Cs should fuse perceptually and the rhythm should be heard clearly.

General Objectives and Hypotheses

One of the objectives of the study was to assess the relative importance of auditory-grouping cues by creating competition among them. Stimulus onset asynchrony (SOA) was expected to have a powerful effect on cross-spectral grouping, because it is a highly reliable cue for the segregation of sound-producing events (see Darwin & Carlyon, 1995). In a natural context, sounds coming from different environmental sources are likely to have some degree of temporal overlap; however, they are unlikely to be perfectly coincident in time. Given the adaptive value of detecting deviations from perfect temporal coincidence, an empirical question of interest was to estimate the physical range of tolerance for the perceived simultaneity of sound events. We expected that the minimum temporal deviation between two tones from perfect synchrony needed to trigger the perception of separate sound events would be in the range of 20 to 40 ms (e.g., Dannenbring & Bregman, 1978; Rasch, 1978). Indeed, Turgeon et al. (2002) obtained estimates in this range for noise bursts using the same RMR paradigm. Let us now turn to the free-field study, looking at how the spatial separation of sound sources interacts with two of the most robust cues for the perceptual fusion of concurrent tones: harmonicity and temporal coincidence.

Experiment 1: Fusion of Complex Tones in a Free Field

Specific Objectives and Hypotheses

We expected simultaneity of onset and offset to make a much greater contribution to the fusion of complex tones than would their harmonic relations or their separation in space. This expectation was based on the high ecological validity of temporal coincidence for the perception of components as a single event as well as the empirical evidence showing its powerful effect on the fusion of components (Bregman & Pinker, 1978; Dannenbring & Bregman, 1978; Darwin & Ciocca, 1992; Darwin & Hukin, 1998; Darwin & Sutherland, 1984; Turgeon et al., 2002). Given the extensive evidence for the role of simple harmonic ratios in promoting perceptual integration (Demany & Semal, 1988; Hartmann et al., 1990; Moore et al., 1986; Roberts & Brunstrom, 1998), fusion at any given SOA should be higher for Ds and Cs sharing the same F0. Therefore, such differences in fusion should be reflected in different SOA thresholds.

Past results suggest that the perceptual organization of sounds is influenced by the spatial separation of their sources (Kidd et al., 1998; Yost, 1997; Yost et al., 1996). However, the results of a prior free-field RMR experiment (see Turgeon et al., 2002, Experiment 2), showed that presenting in different speakers noise-burst Ds and Cs (referred to as "maskers" and "flankers," respectively, in that article) diminished only weakly their fusion, compared with when they came from the same speaker. This can be contrasted with the strong effect of SOAs of 36 and 48 ms, which fully segregated the Ds and Cs at all angular separations of their sources ($\Delta\theta$s) from 0° to 180°. The weak effect of $\Delta\theta$, compared with SOA, is probably related to temporal coincidence acting as a more robust cue than a common location in space for sound-source determination. Spatial information about an acoustic source is often unreliable or ambiguous, owing to the combined effects of diffraction, echoes, and reverberation (Bregman, 1990, pp. 36–38). Therefore, $\Delta\theta$ was expected to have at best a weak effect on cross-spectral fusion.

Method

Participants

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

Stimulus Generation and Presentation

All stimuli were synthesized at a sampling frequency of 20 kHz using MITSYN signal processing software (Henke, 1997) and presented via a 16-bit digital-to-analog converter (Data Translation 2823). Signals were low-pass filtered at 5 kHz through a Rockland Dual Hi/Lo Filter Model 852 using a flat amplitude (Butterworth) response with a 48-dB/octave roll-off. Listeners sat in front of a semicircular array of 13 speakers, each 1 m away from the listener. This arrangement allowed a maximum spatial separation between speakers of 180°. The speaker array was situated in a sound-attenuating chamber of dimensions 2.2 m × 2.2 m × 1.5 m (height). Its walls were covered with Tempest acoustic material and 10-cm-thick Sonex sound-absorbing material to reduce reverberation. The ceiling was made of acoustic panels, and the floor was covered with a thick, absorbent carpet. For the source level used in this experiment, echoes and reverberation were reduced to below the level of the ventilation noise in the chamber (approximately 40-dB sound pressure level [SPL]).

The head of the listener rested on a semicircular support stand, so that the nose was pointing in the direction of the central speaker of the array. Although the stand constrained the possible head movements, it did not ensure that the head of the listener was completely immobile. Accordingly, each listener was explicitly instructed to keep his or her head immobile. The experimenter used a laser beam marker to monitor the position of the listener's head with a high degree of precision. Each four-partial tone was set, at each speaker location, to a level calibrated as equal to that of a 1-kHz tone presented at 66-dB SPL through the central speaker. The intensity was measured at "A" weighting with a General Radio Company sound-pressure meter (Type 1565-B).

The experiment was run online using a MAPLE program (Achim, Ahad, & Bregman, 1995). Because of the constraints of the available space and of keeping the listener's head immobile, the experimenter sat behind the speaker array and recorded the listener's responses, provided verbal feedback on accuracy, and initiated new trials. To facilitate the concentration of the listeners on the sounds, we dimmed the lights and used curtains to cover the front and back of the speakers. The curtains also dampened the propagation of sound behind the speaker array.

Individual Tones of the Sequence

Temporal relations. Each tone was 48-ms long, including 8-ms quarter-sine onset and offset ramps. Figure 2a shows that the Cs could be either simultaneous with the Ds or delayed by a constant SOA. Given the constant 48-ms duration of the D and C tones, SOAs of 0, 12, 24, 36, and 48 ms corresponded to temporal overlaps of 48 (full), 36, 24, 12, and 0 ms (none), respectively. The components of the Ts, Ds, and Cs were all in sine phase. When the Ds and Cs were harmonically related, the positive peaks of their waveforms were aligned at the pitch period throughout their time of overlap. The frequencies, temporal parameters (duration, onset, and offset) and energy (root-mean-square level) were the same for the Ts and the Ds.

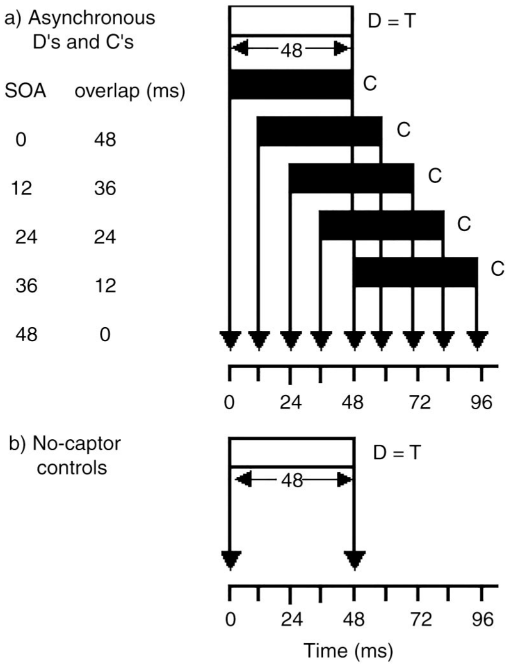

Figure 2. a: The temporal relations between the distractor (D) and captor (C) tones. The C tones could be simultaneous with the D tones or delayed by 12, 24, 36, or 48 ms. The different stimulus onset asynchronies (SOA) corresponded to different durations of temporal overlap between the C and D tones; these are indicated in the left columns. b: The no-C control where only D and acoustically identical target (T) rhythm tones were present.

Spectral relations. There were four possible spectra for the tones: the first four odd or even harmonics of 300 or 333 Hz at equal level. Note that the even-harmonic cases correspond to the first four consecutive harmonics of 600 or 666 Hz. In any condition, the same spectrum was used for every T and D tone in the sequence. Together, the Ts and Ds formed the masked-rhythm sequence, which was presented in isolation for the no-C control conditions (see Figure 2b). In all the other conditions, some C tones were added. When the Ts and Ds were odd harmonics of F0, the Cs were even harmonics and vice versa. The Cs either shared a common F0 with the Ds or not. For each of the four combinations of odd–even and F0 factors, there were two versions of the RMR sequence—distinguished by which of the Ds (and Ts) or the Cs had the higher pitch (even harmonics of F0) and which had the lower pitch (odd harmonics of F0). When Ds and Cs shared a common F0, their nearest harmonics were at least one equivalent rectangular bandwidth apart (Glasberg & Moore, 1990). When Ds and Cs were on different F0s, the higher components (5 and above) would have shown a greater degree of within-band interaction.

Spatial relations. The masked-rhythm sequence (Ts and Ds) and the irregular Cs were either presented in the same central speaker or else in two different speakers equally distant from the central speaker. The speakers could be off center by 30°, 60°, or 90°; these relative positions of the sources of the Ds and Cs yielded three $\Delta\theta$s: 60°, 120°, and 180°.

Structure of the Sequences

Listeners were asked to discriminate between two rhythmic patterns. Each was repeated to form a sequence that had a total duration of 8.8 s, was composed of 15 tones, and had a mean tempo of 1.7 tones per second. The two rhythms were different arrangements of a short (384-ms) and a long (768-ms) intertone interval (ITI). These intervals correspond to 8 and 16 times the tone duration of 48 ms. Rhythm 1 repeated a cycle of short, long, short, and long ITIs three and a half times. Rhythm 2 repeated a cycle of short, long, long, and short ITIs three and a half times. This ensured that both rhythms began and ended with an alternation of short and long ITIs. Although the temporal structure of Rhythm 1 gave rise to perceptual grouping of tones by pairs, the structure of Rhythm 2 gave rise to one in which triplets alternated with a single tone.

The irregularly spaced Ds and Cs were inserted into the sequence during the short and long ITIs so as to ensure no temporal overlap with the Ts. Each D (or DC unit) was fitted into a 96-ms temporal window, which corresponds to the maximum duration of any DC unit, namely, the 48-ms D plus the 48-ms C for the 48-ms SOA condition. Depending on the SOA value, a silence was added to complete the 96-ms total duration. Similarly, there was also a 48-ms silence after each 48-ms D in the no-C controls. Each 192-ms interval (2 × 96 ms) contained one D or DC unit (as appropriate for the condition) with an onset time randomly chosen in the range 0–96 ms within this interval. Therefore, each short and long ISI contained two and four irregular Ds, respectively. The variability in the distribution of irregular intervals was the same in all conditions and was intended to render the timing of successive Cs (when present) useless as a cue to the target rhythm.

To prevent listeners from directing their attention to the beginning or the end of the sequence, we started the irregular tones at a variable time before, and we ended at a variable time after, the regular ones. A random number of 192-ms intervals, from 1 to 11, with one irregular D each, was played before the rhythm began. The remaining number of 192-ms intervals required to complete a total of 12 was played after the rhythm had stopped so that the total duration of the sequence was kept constant across trials.

Design

In terms of spectral relations, eight of the possible pairings of the Ds and Cs were used (for full details, refer back to the section *Individual Tones of the Sequence*). Four of these pairings were harmonically related (same-F0 conditions) and four were not (different-F0 conditions). Each of these combinations was presented at five SOAs (0, 12, 24, 36, or 48 ms) and four $\Delta\theta$s (0°, 60°, 120°, or 180°). For each $\Delta\theta$, there were two presentations, one with the masked rhythm in the left hemifield and one in the right. This gives 320 stimuli ($8 \times 5 \times 4 \times 2$). In addition, no-C controls (64 stimuli) were included as baselines for all combinations of D F0 and angular separation. The purpose of the no-C controls was to verify that the rhythm was camouflaged perceptually by the Ds for each F0 and location. The full set of stimuli was presented in random order across trials.

Procedure

The listeners had to judge which one of the two rhythms was embedded in the sequence and how clearly it was heard on a 5-point scale, where 1 = *guessing*, 2 = *very unclear*, 3 = *unclear*, 4 = *clear*, and 5 = *very clear*. After each trial, verbal feedback about the accuracy of rhythm identification was provided. There was a short training session in which listeners heard a warning tone followed by one of the two rhythms in isolation. They were instructed to direct their attention to the location of the warning tone and to indicate which of the two rhythms was then played. One of the two rhythms was played at one of the speaker locations (chosen randomly on each trial) until the listeners reached the criterion of 10 consecutive correct identifications. This training session was followed by a practice session in which each combination of SOA, $\Delta\theta$, and harmonicity was randomly presented. This session allowed the listeners to become familiar with the task and ensured stable performance during the experimental sessions. During the experiment proper, a 1000-Hz warning tone was played on each trial, just before and in the same speaker as the masked rhythm, so that listeners could direct their attention to the location of the forthcoming rhythm. The head of the listener remained still throughout each block of trials.

RMR as an Estimate of the Degree of Perceptual Fusion in a Free Field

When the Ds and Cs perceptually fuse, they form DC units with changed properties. For instance, when 600-Hz F0 Ds are presented simultaneously with 300-Hz F0 Cs, in speakers separated by 120°, they acquire a 300-Hz pitch and their perceived location is somewhere between the central speaker and the speaker presenting the Cs. These new properties of the DC units allow the listener to hear them as distinct from the rhythmic tones, which are still perceived as coming from the veridical speaker location (60° from center in this example) and having a 600-Hz pitch. The observed fusion of the Ds and Cs was not all or none. Rather, the more the irregular Ds and Cs fused, the clearer the rhythm was perceived.

Results and Discussion

Measures of Sensitivity to the Target Rhythm

For a given condition, the accuracy measure used was the proportion of correct rhythm identifications (P_c). In addition to the objective P_c measure, the rated clarity (RC) of the identified rhythm provides a subjective assessment of the degree of fusion of the D and C tones. Given the single-interval nature of the task, the response-bias indices *c* were computed for the P_c scores (Macmillan & Creelman, 1991). There was no evidence of systematic response bias in any of the experiments reported here and so P_c scores were not converted to detection indices, d'. Instead, the P_c scores were arcsine transformed for the analysis of variance (ANOVA) to account for the compression that occurs as P_c scores approach the ceiling of 1.0 (Howell, 2002, pp. 347–348).

Computation of SOA Thresholds

SOA thresholds were determined for individual psychometric functions to estimate the minimum onset asynchrony leading to the perception of the concurrent Ds and Cs as separate sound events. SOA thresholds were evaluated for eight spectro–spatial relations between the Ds and Cs: common F0 at $\Delta\theta$s of 0°, 60°, 120°, and 180° and their counterparts with different F0s. Each participant's SOA threshold was determined by fitting the psychometric function using a Weibull (1951) function, which minimizes the mean square estimates of error for P_c as a function of SOA. For a given participant, the SOA threshold was estimated as the value yielding a P_c of 0.75. For some participants, the SOA threshold could not be determined from the fitted function, because of either a ceiling performance (all P_cs higher than 0.75) or a below-threshold performance (all P_cs lower than 0.75). Above-threshold performance was observed mainly for the harmonic condition at 0° $\Delta\theta$. Participants whose performance was all above or below threshold in a given condition were not used to estimate the mean SOA threshold for that condition. The majority of participants were included in the mean threshold for each of the eight spectro–spatial conditions (minimum $N = 10$). The mean goodness of fit (r^2) ranged from 0.95 to 0.97 ($p < .05$ in all cases).

Description of the Main Trends

An exploratory four-way ANOVA on the transformed P_c scores showed that the main effect of D F0 and of its interaction with other factors were not significant. Therefore, the transformed P_c scores and the RC scores were collapsed across the four levels of D F0, and three-way within-subjects ANOVAs were performed on the two sets of scores. A conservative Greenhouse-Geisser criterion (Howell, 2002) was used to assess the significance of each effect. Figure 3 displays the P_c scores (top panels) and the RC scores (bottom panels) for the same-F0 condition (left-hand panels) and for the different-F0 condition (right-hand panels).

No-C controls. The top panels of Figure 3 indicate that the four no-C controls yielded mean P_c scores between 0.53 and 0.56. Intersubject standard errors (*SE*s) were between 0.03 and 0.05. These values did not differ significantly from chance, as would be expected if the rhythm was perceptually masked in the absence of Cs. The P_c scores are consistent with the RC ones, shown in the bottom panels of the figure; these are close to 1, which corresponds to *guessing*.

Effect of SOA and temporal limits for event segregation. The temporally coincident Ds and Cs strongly fused perceptually, as indicated by the high P_c and RC scores at 0-ms SOA for both same- and different-F0 conditions. At 0-ms SOA, the mean RC scores were all close to the maximum RC of 5.0 (very clear) and the P_c scores were all at ceiling (i.e., 0.99 or 1.00). This was true for all $\Delta\theta$s tested. This suggests that onset synchrony caused frequency components to be perceptually fused, regardless of whether they were harmonically related and regardless of whether they came from the same or different locations. The top panels of Figure 3 show a clear monotonic decrease in P_c as a function of SOA, $F(4, 32) = 81.24$, $p < .001$. The bottom panels of the figure show that this also holds for RC scores, $F(4, 32) = 143.11$, $p < .001$. This robust effect of SOA on the fusion of the Ds and Cs is consistent with the results for the fusion of brief noise bursts in a free field (Turgeon et al., 2002). However, it is noteworthy that two-tailed *t* tests showed that P_c scores for an SOA of 48 ms were significantly greater than for the no-C control, whether the Ds and Cs shared the same F0, $t(17) = 2.66$, $p < .05$, or not, $t(17) = 3.41$,

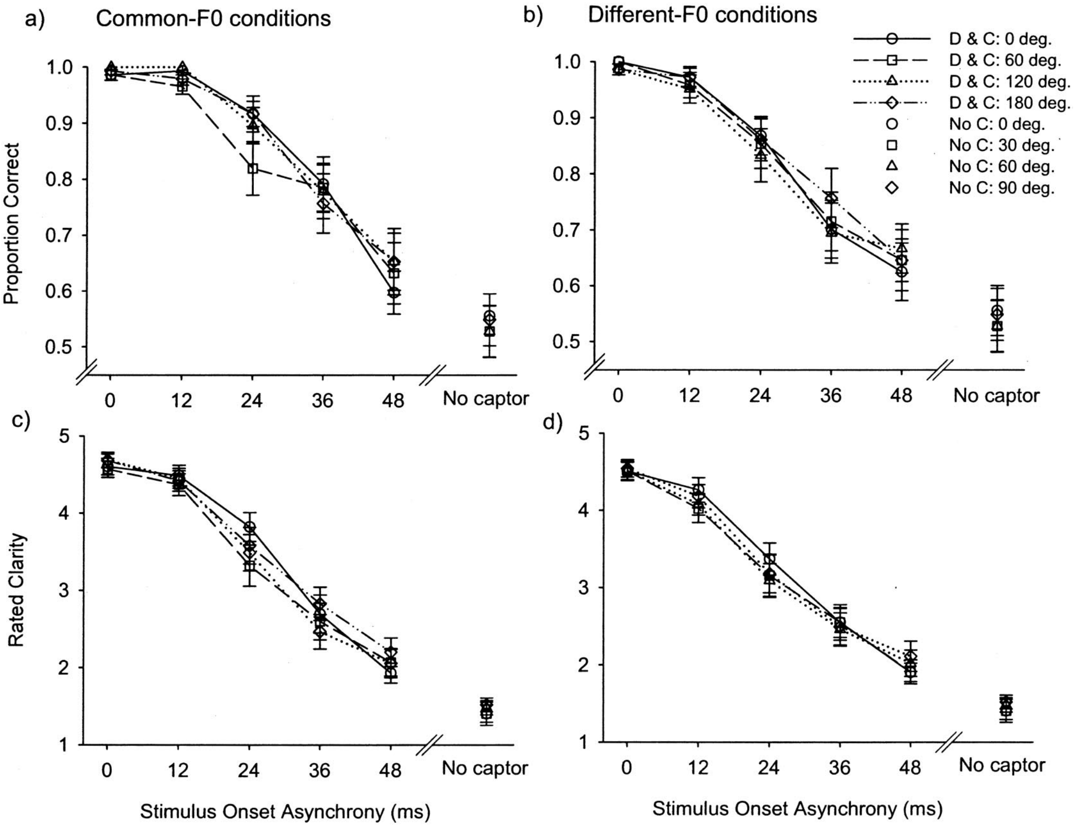

Figure 3. The mean proportion correct (P_c) of rhythm identifications (a and b) and rated clarity scores (c and d) are shown with intersubject standard errors ($N = 18$). Both measures decrease monotonically as a function of stimulus onset asynchrony, whether the distractor (D) and captor (C) tones share a common fundamental frequency (F0; a and c) or not (b and d). The degree of fusion between the D and C tones appears to be independent of the angular separation of the sound sources, whether it is assessed through P_c or rated clarity scores. The right-most symbols indicate the near-chance-level performance for each of the angular positions of the no-C controls (i.e., masked-rhythm only). These served as baselines for the corresponding conditions with Cs. The angular separation values quoted for the no-C controls are relative to 0° azimuth, and hence half the value of their counterparts, for which the Cs were presented at an equal distance from the center as the masked rhythm in the other hemifield. Error bars represent intersubject standard errors.

$p < .01$. This above-chance performance is surprising given that there is no temporal overlap between Ds and Cs at this SOA. This finding is discussed further in the discussion of Experiment 3.

Figure 4 shows the mean SOA thresholds estimated for the eight spectro–spatial conditions. These results suggest that an SOA of about 30–40 ms triggers the perception of brief tones as separate events. This finding agrees with the literature on auditory grouping, reviewed by Darwin and Carlyon (1995), showing that an SOA of 30–40 ms is required for removing a partial from contributing to the overall timbre and lateralization of a complex sound. A similar SOA is needed to hear two events in a mixture that would sound unified in the absence of an SOA (Dannenbring & Bregman, 1978; Rasch, 1978). In contrast, the SOA needed to prevent a partial from influencing the computation of the low pitch of a complex tone is about 300 ms (Darwin & Ciocca, 1992). This discrepancy between the temporal limits for pitch and event perception may be related to differences in their underlying neural mechanisms (Brunstrom & Roberts, 2000).

Effect of spectral and spatial factors. Overall, the P_c and RC scores tended to be slightly higher for Ds and Cs that shared a common F0 (Figures 3a and 3c) than for those that did not (Figures 3b and 3d). It is the consistency of the harmonicity effect across individual listeners for SOA values of 12, 24, and 36 ms that probably accounts for its statistical significance for both P_c scores, $F(1, 17) = 6.22$, $p < .05$, and RC scores, $F(1, 17) = 69.21$, $p < .001$. For P_c scores, the effect of harmonicity depended on the SOA value, the effect being absent when there was either a full overlap or no overlap. This is shown by the significant two-way interaction between SOA and harmonicity on RC scores, $F(3, 55) = 3.97$, $p < .05$; note also that this interaction term approaches significance for the transformed P_c scores, $F(3, 54) = 2.33$, $p = .08$. The absence of any harmonicity effect at 0-ms SOA is probably due to the near-ceiling performance, and the lack of such an effect at 48-ms SOA is unsurprising given that the Ds and Cs do not temporally overlap.

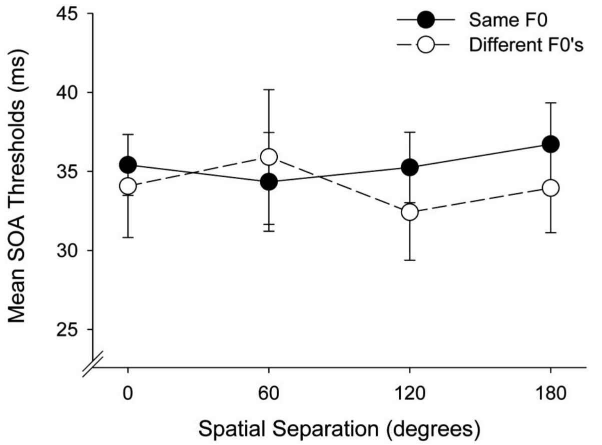

Figure 4. Mean stimulus onset asynchrony (SOA) thresholds, with intersubject standard errors (shown by the error bars), as a function of the angular separation between the distractors (and target rhythm) and the captors, when they have a common fundamental frequency (F0; filled circles) and when they do not (open circles). Individual SOA thresholds were determined from the 0.75 proportion-correct level of the best-fitting Weibull psychometric function for each condition.

There was no main effect of source separation on rhythm discrimination, $F(2, 39) = 0.69$, $p > .50$, although there was a small but significant effect on the subjective RC scores, $F(2, 29) = 3.64$, $p < .05$. There was also a small but significant interaction between SOA and $\Delta\theta$ for the RC scores, $F(6, 99) = 3.27$, $p < .01$. Figure 4 shows that the mean SOA thresholds were highly consistent across the different spatial and spectral relations of the Ds and Cs. The mean thresholds all fell between 32.42 ms (for $\Delta\theta$ of 120° and different F0s) and 36.72 ms (for $\Delta\theta$ of 180° and the same F0). Harmonicity and spatial separation thus seem to affect only slightly the temporal disparity needed for the perception of separate events in the RMR paradigm.

It is worth noting that for Ds sharing a common F0 with the Cs, there was very little difference in P_c scores between the low-pitch Ds (i.e., odd harmonics of 300- or 333-Hz F0) and their high-pitch counterparts (i.e., even harmonics of 300- or 333-Hz F0). The Ts and DC units can be differentiated by pitch and timbre in the latter configuration but only by timbre in the former. Clearly, a difference in low pitch is not required to segregate Ts and DC units—a difference in timbre is quite sufficient.

Summary of Results

Temporal coincidence, and deviations from it induced by onset asynchrony, was by far the most important factor for the perception of short-duration tones (DC units) as one or two sounds. Whereas Ds and Cs fused into a single DC event when they were synchronous, they were segregated as two distinct events when they were separated by an SOA of about 35 ms. Strong perceptual fusion was clearly shown by the near-perfect rhythm identification at 0-ms SOA. On the other hand, clear segregation was shown by worse performance at 48-ms SOA. Intermediate values of SOA of 12, 24, and 36 ms produced intermediate degrees of grouping in which the Ds and Cs were neither fully fused nor fully segregated.

Despite the ecological validity of simple harmonic ratios for sound-source determination, this factor only contributed weakly to the fusion of spectrally and spatially distributed brief tones at intermediate asynchronies in the present study. Nonetheless, the greater fusion induced by harmonicity (same-F0 conditions vs. different-F0 conditions) was significant, despite the fact that DC units in the different-F0 conditions would have involved greater within-band interactions. These interactions would have made the components of the Cs and Ds less distinguishable from one another, hence overestimating the degree of fusion. Furthermore, the short duration (48 ms) of the tones and the even shorter periods of overlap between the Cs and Ds (from 48 to 0 ms) probably contributed to the weak effect of harmonicity on the observed perceptual fusion. Indeed, there is evidence that the perception of short complex tones as unified tolerates greater departures from harmonic relations among their partials than for otherwise equivalent tones of longer duration (e.g., Moore, 1987).

The angular separation of the sources did not affect fusion at all, regardless of whether it was estimated from P_c scores or from SOA thresholds. Overall, the present results suggest a limited or absent contribution of sound-source separation in space to sound-source segregation over time, at least when sources are concurrently active. This finding is consistent with earlier studies indicating that sound localization can be a weak cue for influencing simultaneous grouping (e.g., Broadbent & Ladefoged, 1957; Darwin, 1997).

Experiment 2 explored whether dichotic separation would be more efficient for sound segregation, as it represents the extreme case of interaural difference.

Experiment 2: Fusion of Complex Tones Under Diotic and Dichotic Presentation

Objectives and Hypotheses

This experiment was intended to maximize any possible effects of spatial separation by presenting the masked rhythm (T and D tones) and the C tones in opposite ears. By using otherwise identical complex tones to those used in Experiment 1, we also hoped to generalize the observed effects of SOA and harmonicity on fusion in a free-field context to fusion under headphone presentation. Dichotic presentation was expected to result in significantly lower binaural fusion relative to diotic presentation. Indeed, Turgeon et al. (2002) showed in a prior RMR study that dichotic presentation reduces the fusion of noise bursts by more than the effect of source separation in a free field. Given that the paradigm, task, stimulus structure, computation of the results, and statistical analyses are essentially the same as those of Experiment 1, the *Method* and *Results and Discussion* sections are restricted to identifying what is different in Experiment 2.

Method

The stimuli were presented through Sony NR-V7 headphones at 67-dB SPL ("A" weighting) to listeners seated in a sound-attenuating chamber (Industrial Acoustics, 1202). Listeners initiated new trials and recorded their responses directly via the computer and were provided with on-screen feedback about the accuracy of rhythm identification at the end of each trial.

The masked-rhythm sequence (Ts and Ds) and the irregular Cs were either presented in both ears or in separate ears. Diotic and dichotic conditions were presented in separate blocks of trials, with the order counterbalanced across participants; all other factors were randomized within each block. For the dichotic presentation, the masked rhythm was presented in the same ear (left or right) throughout a block of trials. The Cs, when present, were presented in the contralateral ear. Note that the dichotic no-C controls are, in effect, monaural presentations. There were 4 replications for each combination of DC pairing (8 levels) × SOA (5 levels) × spatial separation (2 levels), giving a total of 320 stimuli. There were also 4 replications for each no-C control (64 stimuli).

Results and Discussion

Description of the Main Trends

As for Experiment 1, the transformed P_c scores and the RC scores were analyzed using three-way within-subjects ANOVAs. Figure 5a displays the P_c scores and Figure 5b displays the RC scores for the conditions tested, collapsed over the four levels of D F0. The diotic and dichotic (monaural) no-C controls yielded mean P_c scores of 0.53 and 0.56 (intersubject *SE*s of 0.03 for both) and mean RC scores of 2.32 (*SE* = 0.32) and 1.69 (*SE* = 0.20), respectively. These values are shown by the isolated bold symbols in Figure 5. Although listeners did not consistently report guessing the target rhythm, the near-chance rhythm-identification performance obtained for the no-C controls verified that the rhythm was perceptually masked in the absence of Cs.

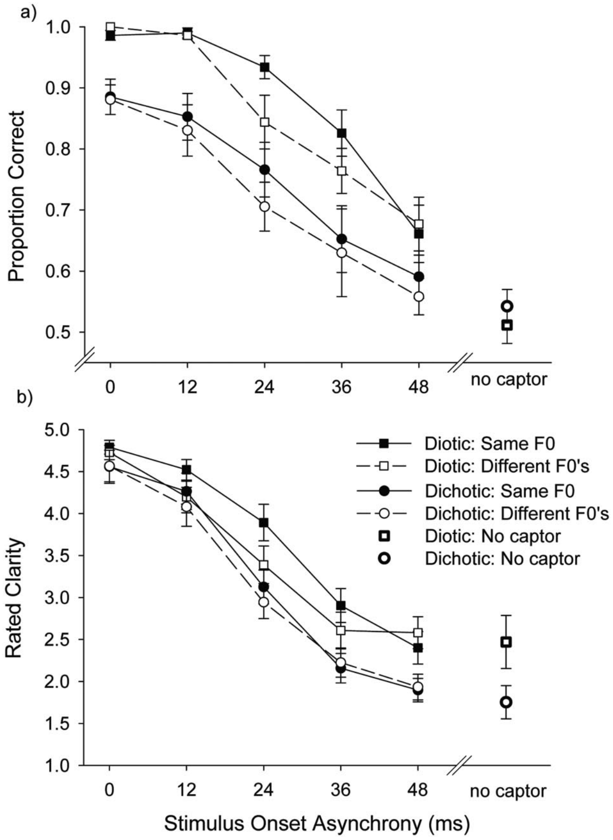

Figure 5. The mean proportion correct of rhythm identifications (a) and rated clarity scores (b) are shown with intersubject standard errors (denoted by the error bars; *N* = 18). Both measures decrease monotonically as a function of stimulus onset asynchrony. The scores are consistently higher (i.e., more fusion) for the diotic than the dichotic presentation of the distractor and captor tones (squares vs. circles). Scores also tend to be higher when the distractors and captors share a common fundamental frequency (F0) than when they do not (filled vs. open symbols). The diotic (bold square) and monaural (bold circle) no-captor controls served as baselines for the diotic and dichotic conditions, respectively. Performance for these controls is close to chance.

Figure 5 shows a clear monotonic decrease in P_c scores with SOA in all conditions, $F(2, 40) = 68.96, p < .001$. The same holds for RC scores, $F(2, 30) = 35.73, p < .001$. From the 0-ms to the 48-ms SOA, the mean P_c scores decreased from 0.88 (at least) to 0.68 (at most), and the RC scores decreased from 4.5 (at least) to 2.6 (at most). These results suggest that asynchrony decreases the fusion of brief tones across the spectrum and across ears. The dichotic conditions (symbols = circles) yielded smaller mean P_c scores than their diotic counterparts (symbols = squares), $F(1, 17) = 27.55, p < .001$, particularly at 0-ms, $t(17) = 2.80, p < .05$, and 12-ms SOA, $t(17) = 2.84, p < .05$. Of course, one might expect to observe the greatest effect of the diotic–dichotic distinction when there is most temporal overlap between the Ds and Cs.

The P_c scores show that the Ds and Cs with a common F0 (filled symbols) tended to fuse more than those with a different F0 (open symbols), $F(1, 17) = 8.55, p < .01$. That the effect of spectral relations depended on the magnitude of the SOA is reflected in the two-way interaction between those factors for both P_c scores, $F(3, 54) = 3.66, p < .05$, and RC scores, $F(2, 40) = 11.48, p < .001$. The absence of a common-F0 effect at 0 and 12 ms SOA is probably due to the near-ceiling performance at those SOAs. Its absence is expected at 48-ms SOA, for which there is no overlap of the Ds and Cs. Clear and substantial differences in P_c and RC scores due to spectral relations were restricted to only two conditions: the 24-ms and the 36-ms SOA diotic stimuli.

Figure 6 shows the mean SOA thresholds for the four spectro–spatial conditions. As in Experiment 1, participants whose performance was all above or below the 0.75 P_c threshold for a given condition were not used to estimate its mean SOA threshold. The mean thresholds for the two diotic and the two dichotic conditions were based on 15 and 17 of the 18 listeners, respectively. The mean goodness of fit (r^2) was .97 or above ($p < .05$ in all cases). The thresholds were clearly lower in dichotic (two right-hand bars) than in diotic presentation (two left-hand bars). This difference is larger when Ds and Cs share a common F0 (13.8 ms, as shown by the two filled bars) than when they do not (8.7 ms, as shown by the two open bars). Overall, these results suggest that an SOA of about 30 to 40 ms is required for the segregation of short-duration tones (e.g., the perception of Ds and Cs as distinct events). This range corresponds to the lead time required for a partial to no longer contribute to the timbre of a complex tone (e.g., Bregman & Pinker, 1978) or its lateralization (e.g., Hill & Darwin, 1993). Although the asynchrony needed for segregating temporally contiguous brief sounds can be lowered by presenting them in separate ears, it does not seem to be much affected by whether they share a common F0.

Summary of Results and Interim Discussion

The effect of spatial separation by ear of presentation, though weaker than the effect of SOA, is robust. It can lower substantially the temporal disparity needed to segregate a sound from another by as much as 12 ms, regardless of whether they are harmonically related. Its effect on cross-spectral segregation holds across all SOA values from 0 to 48 ms, being strongest at intermediate values of SOA (12, 24, and 36 ms). However, dichotic stimulation is not typical of everyday listening environments, except where different sound sources are very close to opposite ears (e.g., when someone is whispering in one ear while an insect is buzzing in the other). For this reason, the separation of sound sources in a free field is considered to be more representative of the true contribution of spatial separation to sound-source segregation. The results of Experiment 1 indicate that spatial separation in a free field has little or no effect on the fusion of brief tones. Experiment 2 replicated the weak effect of harmonicity and the strong effect of asynchrony on the grouping of multipartial tones of short duration. Though small in magnitude, the effect of harmonicity in promoting fusion appears to be stronger when the components share a common lateralization (here, through diotic presentation).

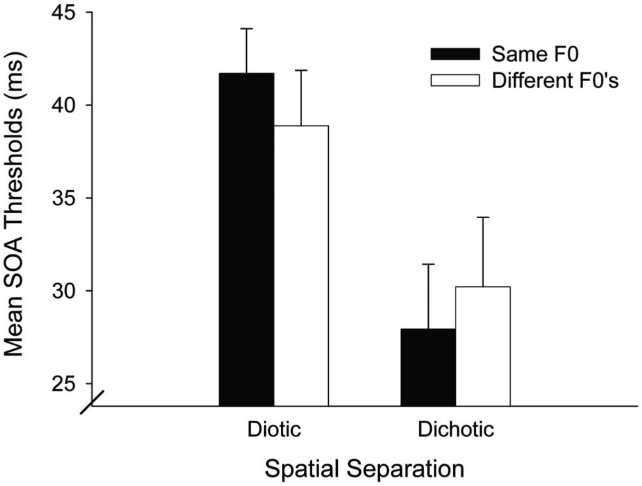

Figure 6. Mean stimulus onset asynchrony (SOA) thresholds, with intersubject standard errors (denoted by the error bars), are shown for diotically and dichotically presented distractor and captor tones sharing a common fundamental frequency (F0; filled bars) or not (open bars). Individual SOA thresholds were determined from the 0.75 proportion-correct level of the best-fitting Weibull psychometric function for each condition.

One interpretation of the results of Experiments 1 and 2 is that asynchrony, especially onset asynchrony, dominates the grouping of short-duration tones. This interpretation was proposed in a prior RMR study (Turgeon et al., 2002). However, it would be premature to accept this proposal at this stage. This is because the experiments reported so far used a stimulus design that confounds the factors of onset asynchrony and degree of temporal overlap. Perhaps the decline in the accuracy of rhythm identification with increasing SOA reflects the increasing exposure of the Ds. Indeed, an SOA of 48 ms gives rise to DC units in which the Ds and Cs do not overlap at all. Previous studies of the effects of onset asynchrony on auditory grouping are often similarly confounded (e.g., Bregman & Pinker, 1978). Furthermore, the increasingly brief period of overlap for Ds and Cs as the SOA is increased may have reduced the scope for factors like harmonicity and spatial location to influence performance. Finally, reducing the increase in within-channel interactions among the partials of the Ds and Cs when they do not share a common F0 may reveal a greater effect of harmonicity than observed so far. Experiment 3 attempted to overcome these limitations.

Experiment 3: Relative Contribution of Asynchrony and Temporal Overlap

Objectives and Hypotheses

Experiment 3 attempted to tease apart the relative contributions of onset asynchrony and temporal overlap by using Ds and Cs that vary in their degree of onset asynchrony but which maintain a constant duration of temporal overlap. Specifically, the SOA magnitude was the same as that used in Experiments 1 and 2. However, unlike the previous stimuli, the Ds and Cs used in Experiment 3 had a constant period of overlap, corresponding to the full duration of the Ds (48 ms). This approach ensured that no portion of the Ds was heard in isolation when the Cs were present. If SOA has a similar effect on rhythm identification under these conditions, then one can be confident that SOA per se, rather than degree of temporal overlap, was the key factor affecting performance in Experiments 1 and 2. In contrast, a greatly reduced effect of SOA would suggest that earlier studies had overestimated the importance of onset asynchrony in determining the perceptual fusion of brief tones and would also demonstrate a major role for degree of temporal overlap in governing performance.

Another goal of Experiment 3 was to optimize the measurement of the effect of harmonic relations on the fusion of brief tones. This was done in two ways. First, by maintaining the duration of temporal overlap of the Ds constant at 48 ms for all SOAs tested, the auditory system should have a better sampling of the fundamental period. Second, the different-F0 condition was replaced by a shifted-C condition in which harmonic relations between the Cs and Ds were disrupted not by changing the F0 of the Cs but instead by adding a fixed increment to the frequency of each component of the Cs. This manipulation, known as frequency shifting, has previously been used in studies of low pitch (e.g., de Boer, 1976; Patterson, 1973) and of spectral grouping (e.g., Roberts & Brunstrom, 1998, 2001). The advantage of using frequency-shifted Cs with the same nominal F0 as the Ds is that the component spacing of the Cs does not change on a linear scale. Therefore, the higher components will not come into such close proximity in the combined DC units as would be the case if the Cs and Ds differed in F0 by 10% (as in Experiments 1 and 2). This reduces the potentially confounding effect of differences in the extent of within-channel interactions between conditions.

Method

The listeners were 7 normal-hearing adults, mostly students at the University of Birmingham who were naive to the purpose of the experiment. The stimuli were presented using a Turtle Beach Montego A3DX Stream sound card through Sennheiser HD480-13II headphones at 67-dB SPL (“A” weighting). No external filtering was required. All other features of the stimulus generation and presentation were the same as in Experiment 2.

Stimuli

The temporal structure of the two rhythmic patterns, the temporal density of the Ds, and the variability in the distribution of the irregular intervals between them were the same as in Experiments 1 and 2. Only the duration of the Cs was changed to ensure that temporal overlap with Ds was maintained constant as SOA was changed. Figure 7 shows the temporal relations between the irregular Ds and Cs. The onset asynchrony between the Ds and Cs was varied in the same way as in Experiments 1 and 2 (i.e., SOAs of 0, 12, 24, 36, or 48 ms). For most conditions, the Cs at 96 ms were twice as long as the Ds and rhythmic Ts (48 ms), and the duration of temporal overlap was maintained constant at 48 ms (see Figure 7a). These conditions were to allow the contribution of SOA to be assessed unconfounded by changes in temporal overlap. In addition, a no-overlap control condition was added (see Figure 7b), which was identical to the 48-ms SOA condition of Experiments 1 and 2. As in the prior RMR experiments, there was also a no-C control (see Figure 7c) to ensure that the rhythm was perceptually masked in the absence of Cs.

All stimuli were presented diotically. The Ds and rhythmic Ts were always the first four even harmonics of 300 Hz (corresponding to the first four harmonics of 600 Hz). The Cs were either the first four odd harmonics of 300 Hz or these components were frequency shifted downward by 20% of F0 (i.e., by 60 Hz). This manipulation destroyed the harmonic relations between the Ds and Cs without changing the spacing in Hz between the components of the Cs. This ensured that all components in the DC units were separated by at least one equivalent rectangular bandwidth in both the same-F0 and shifted-C conditions. Frequency shifting the C has relatively little effect on its own perceptual coherence (Roberts & Brunstrom, 1998, 2001).

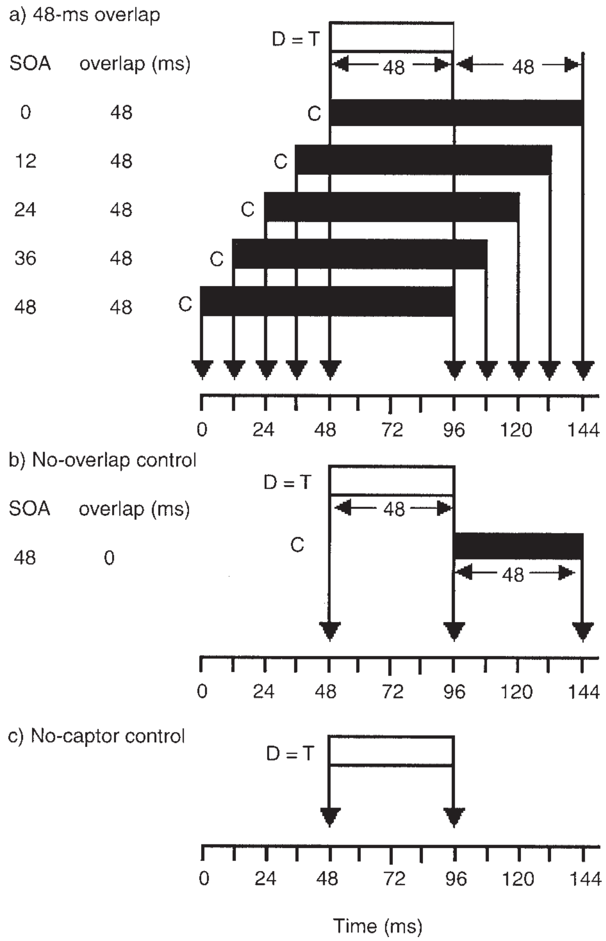

Figure 7. a: The temporal relations between the distractor (D) and captor (C) tones: the 96-ms C tones either began simultaneously with the 48-ms D tones or were leading by 12, 24, 36, or 48 ms. The different stimulus onset asynchronies (SOA) had the same 48-ms duration of temporal overlap of Cs and Ds. b: The 48-ms SOA condition with nonoverlapping C and D tones; note that the C tones are now 48-ms long. c: The no-C control where only D and acoustically identical target (T) rhythm tones were present. As in Experiments 1 and 2, the onset time of each D (or DC unit) was selected randomly over a 96-ms range.

Design

For the conditions with temporally overlapping DC units, the Cs were (a) either harmonically related to the Ds or not and (b) either shared a common onset time with the Ds or were advanced by 12, 24, 36, or 48 ms. The P_c (arcsine transformed) and RC scores obtained were analyzed using two-way within-subjects ANOVAs, with harmonicity (two levels) and SOA (five levels). The no-overlap controls (Cs duration = 48 ms instead of 96 ms) and the no-C controls did not form an orthogonal design with the overlap conditions. Accordingly, *t* tests were performed to compare (a) the overlapping and nonoverlapping 48-ms SOA conditions and (b) the latter conditions with the no-C control. There were 16 replications for each condition, giving a total of 192 stimuli, including the controls. The stimuli were presented in random order across trials.

Results and Discussion

Effect of SOA, Not Confounded by Temporal Overlap

Figure 8 shows that the temporally coincident Ds and Cs strongly fused perceptually, as indicated by the high P_c (see Figure 8a) and RC (see Figure 8b) scores at 0-ms SOA (the left-most data points). For both the same-F0 and the shifted-C conditions, at 0-ms SOA, the P_c scores were at ceiling and the RC scores fell between RCs of 4 (*clear*) and 5 (*very clear*). These results are consistent with the conclusions drawn from Experiments 1 and 2 that temporal coincidence fuses the brief C and D tones, regardless of whether they are related by simple harmonic ratios.

For both same-F0 (filled squares) and shifted-C conditions (open squares), performance remained at ceiling for SOAs of 12 and 24 ms. Performance began to fall for an SOA of 36 ms in the shifted-C condition, but an SOA of 48 ms was needed in the same-F0 condition before it fell below ceiling. SOA thresholds were not computed in this experiment, because the mean performance did not fall below the 0.75 P_c threshold level even at the largest SOA (48 ms). That in itself shows that onset asynchrony has a weaker effect on the fusion of Ds and Cs when temporal overlap is maintained throughout the 48-ms duration of the Ds. It appears that longer Ds, allowing for larger SOAs between the Ds and Cs while maintaining their 100% overlap, would be needed to establish the duration at which the effect of SOA approaches asymptote.

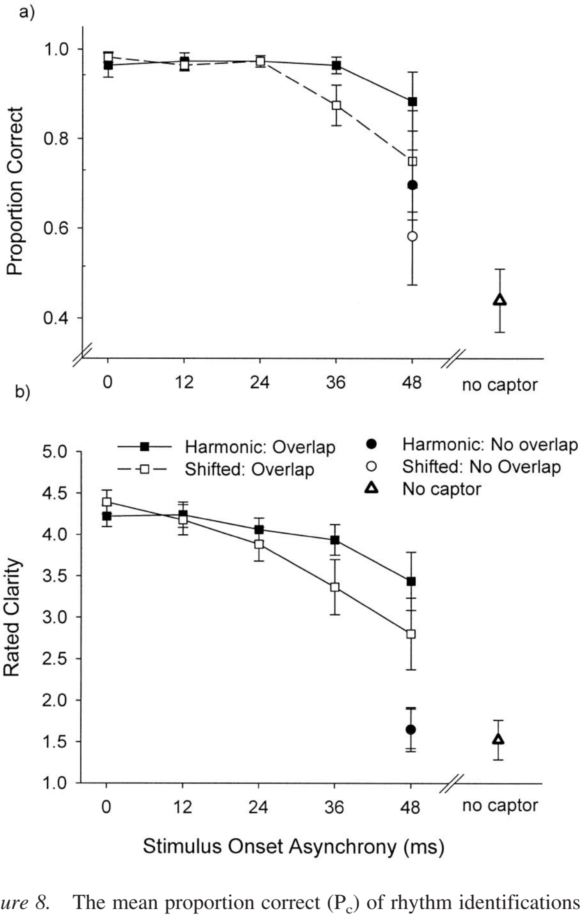

Figure 8. The mean proportion correct (P_c) of rhythm identifications (a) and rated clarity (RC) scores (b) are shown, with intersubject standard errors (denoted by the error bars; $N = 7$), as a function of stimulus onset asynchrony (SOA). When the distractors (Ds) and captors (Cs) are harmonically related (filled symbols), their fusion is less affected by their SOA when their temporal overlap is maintained constant, as shown by the consistently high P_c and RC scores across SOA values. When the Cs are frequency shifted, and so not harmonically related to the Ds (open symbols), SOA has more effect on fusion. However, SOAs greater than 24 ms are needed to bring P_c scores below ceiling. The no-C control (triangles) gave a mean P_c score close to chance and a mean RC score close to 1 (*guessing*).

Statistical analysis of the data using a standard criterion is consistent with the observed weak effect of SOA on both the transformed P_c scores, $F(4, 24) = 3.66, p < .05$, and the RC scores, $F(4, 24) = 9.98, p < .01$. Note, however, that the effect no longer reaches significance for the P_c scores if assessed using the conservative Greenhouse-Geisser correction, as used in Experiments 1 and 2. Although the interaction of SOA × Harmonicity was significant for the RC scores, $F(4, 24) = 5.86, p < .05$, it was not for the P_c scores, $F(4, 24) = 1.50, p > .20$.

Effect of Temporal Overlap

A comparison between the 48-ms SOA fully overlapped conditions and their no-overlap counterparts (squares vs. circles; see Figure 8) can show whether overlap of Ds and Cs is critical in determining their fusion into DC units, thereby releasing the rhythm from masking. Compare the P_c score of 0.88 ($SE = 0.06$) for harmonically related overlapping Ds and Cs against the much lower one of 0.67 ($SE = 0.07$) for their nonoverlapping counterparts; this yields a significant difference between their transformed P_c scores, $t(6) = 2.78, p < .05$. Though there is a trend in the same direction, this effect of temporal overlap does not hold in conditions in which the Cs are frequency shifted away from simple harmonic relations with the Ds, $t(6) = 1.83, p > .10$. The difference in P_c scores of about 0.2 between the overlapping and nonoverlapping 48-ms SOA conditions is twice the magnitude of that found between the smallest (0 ms) and largest (48 ms) SOAs for temporally overlapping harmonic Cs and Ds. For an SOA of 48 ms, the trend toward stronger fusion of overlapping Cs and Ds that share a common F0 may be attributed to the 48-ms temporal window for sampling the pitch period (14.4 repetitions of 3.33 ms).

A Role for Offset Asynchrony or Attention-Based Selection?

The relatively weak effect of SOA on RMR performance when the Ds remain fully overlapped with the Cs is consistent with the finding of Dannenbring and Bregman (1978) that asynchrony assists the segregation of a component into a sequential stream only if it exposes part of that component. However, there are two reasons why Experiment 3 may have underestimated the effect of SOA on the fusion of the Cs and Ds. First, the design did not control for any possible effects of offset asynchrony. Second, the ear had no opportunity to sample the properties of the Ds uncontaminated by those of the Cs. These points are considered in turn.

At least for brief tones, there is evidence that synchrony of offset, as well as onset, facilitates the perceptual fusion of partials (see Darwin & Carlyon, 1995). In Experiment 3, SOA covaried with degree of offset asynchrony, because the effect of SOA was measured while maintaining a constant duration for the Cs and Ds and for their overlap. In particular, the 48-ms SOA overlapping conditions were associated with synchronous offsets of the Ds and Cs that may have encouraged their fusion and hence led to improved RMR performance. Although an effect of offset asynchrony cannot be ruled out, we do not believe the size of the effect to be large for the following reasons. First, offset asynchrony is less than half as effective as onset asynchrony, at least for reducing the contribution of a harmonic to vowel timbre (Darwin & Sutherland, 1984; Roberts & Moore, 1991). Second, an SOA of 0 ms is associated with an offset asynchrony of 48 ms, yet performance was at ceiling in both same-F0 and shifted-C conditions. Clearly, an offset asynchrony alone is insufficient to reduce performance.

Although the Ds were acoustically identical to the rhythmic target tones (Ts), they may not have been perceptually identical as a result of influence by the Cs. For instance, the 48-ms SOA fully overlapping Cs could have affected the overall perceived loudness and/or roughness of the sequence of Ds, even in the absence of perceptual fusion between the Cs and Ds. This would have created loudness and/or timbre cues to distinguish the Ts and Ds, allowing attention-based selection of the regular Ts from the masked rhythm.

Effect of Harmonicity Under Conditions of Constant Overlap

Although the main effect of harmonicity on RC scores approached significance, $F(1, 6) = 5.52$, $p = .057$, this effect was not significant for the transformed P_c scores, $F(1, 6) = 1.24$, $p > .20$. This is probably due to near-ceiling performance at small SOAs. Furthermore, Rasch (1978) reported that harmonic relations had little or no effect on the segregation of asynchronous tones. However, there was a trend toward poorer accuracy for the shifted-C condition at longer SOAs (differences in P_c scores of 0.09 and 0.13 at SOAs of 36 ms and 48 ms, respectively). These differences are actually larger in magnitude than the corresponding differences in P_c scores between same-F0 and different-F0 stimuli under diotic presentation found in Experiment 2 (0.06 at 0.36-ms SOA and 0.02 at 48-ms SOA; see Figure 5a). Care must be taken when comparing the results of Experiments 2 and 3, because different listeners took part and different strategies were used to control the harmonicity of the DC units (different-F0 conditions vs. shifted-C conditions). Nonetheless, this observation is consistent with the hypothesized greater contribution of simple harmonic ratios to perceptual fusion when there is greater temporal overlap and when differences in within-channel interactions are minimized.

Even Nonoverlapping Cs Affect RMR

Figure 8a shows mean P_c scores of 0.67 and 0.56 for the same-F0 (filled circle) and shifted-C (open circle) no-overlap conditions, respectively, and a mean P_c score of 0.43 for the no-C control (bold triangle). Both means are above chance, and the difference between the harmonic no-overlap condition and the no-C control approaches significance, $t(6) = 2.42$, $p = .052$. This finding suggests that the mere presence of the Cs as part of the overall sequence can assist rhythm-identification performance, if the Cs and Ds share a common F0. A similar result has already been noted for the corresponding nonoverlapping conditions in Experiments 1 and 2 (SOA = 48 ms). What might account for the observed above-chance performance when the Cs and Ds do not overlap?

The nonoverlapping stimuli used are all structured such that each D is followed immediately by a C. The absence of a silent interval between them offers the opportunity for several possible sequential factors to operate. These possibilities are evaluated in turn.

1. The Cs may have weakened the representation of the immediately preceding Ds through a process akin to backward recognition masking (Massaro, 1975). The finding of above-chance performance (d' scores around 1) for the corresponding nonoverlapping conditions in the RMR study of Turgeon et al. (2002, Experiment 2, Figure 7b, p. 1826) is consistent with this interpretation. Their study used sequences of narrow-band noises in which the Ds and Cs occupied different spectral regions, whereas the current study used complex-tone Ds and Cs whose components were interleaved. Given that backward recognition masking is not greatly dependent on frequency separation, one would expect above-chance performance for nonoverlapping conditions in both studies, despite differences in the spectral proximity of Cs and Ds.
2. The Cs may have acted as some form of temporal marker for the Ds, allowing attentional processes to identify them and hence to ignore them, perhaps through some inhibitory process. However, there is no evidence that the act of selecting one set of stimuli (the Ds) from the masked rhythm will lead to improved selection of another set (the target rhythm). Indeed, Bregman (1990, p. 452) argued that attentional selection of one set of elements does not create a residual set that is either diminished or organized.
3. In music, a short note can be embellished by following it with a short grace note in close succession. The perceptual unit formed by this ornamentation sounds longer and has a slightly different quality from the original note. A similar phenomenon may occur in our nonoverlapping conditions, such that the Cs act as grace notes for the Ds and so introduce timbre and duration cues to distinguish them from the target rhythm.
4. It is interesting that the clearest case of superior RMR performance when Ds and Cs share a common F0 than when they do not occurs for Ds and Cs that do not overlap in time. It is not clear why this should be the case, but we offer two (possibly related) speculations. First, there is evidence that nonoverlapping harmonically related components can integrate as if they were presented simultaneously under certain conditions. In particular, Ciocca and Darwin (1999) have shown that a low pitch can be generated from a series of nonsimultaneous partials related by simple harmonic ratios. Second, it may be the case that the effectiveness of a grace note is enhanced if it is harmonically related to the original note. Either (or both) of these factors could act to improve rhythm identification.

All of the experiments reported here (except for the temporally overlapping conditions of Experiment 3, shown in Figure 7a), and the second experiment of Turgeon et al. (2002), created SOAs by delaying the Cs relative to the Ds. However, the first experiment of Turgeon et al. differed in that SOAs were created either by delaying or advancing the Cs (with equal likelihood in a given trial). Their results (see Figure 3b, p. 1823) indicated that d' values fell to around chance for the nonoverlapping conditions (SOA = 48 ms). This finding is consistent with the hypothesized contributions of the factors listed above. In particular, a C cannot act as a backward masker or as a grace note for a D if it precedes it.

One other possibility that cannot be ruled out completely is that listeners might learn, through repeated exposure, that there are differences in the possible patterns of Cs associated with the two target sequences, despite the substantial randomization of onset times that was used. If so, there may be some residual information in the timing of the Cs that can act as a weak cue to the identity of the target sequence, even when the Ds and Cs do not overlap. However, the near-chance performance observed by Turgeon et al. (2002, Experiment 1) for their no-overlap conditions suggests that such a learning effect is negligible, because one would not expect it to depend on whether the Cs are delayed or advanced, relative to the Ds.

Summary of Results

SOA continues to influence the accuracy of rhythm identification when the Ds are heard accompanied by the Cs for their entire duration. However, accuracy remains quite high even for an SOA of 48 ms, especially when the Ds and Cs share a common F0. This finding emphasizes the importance of variations in degree of temporal overlap for the results obtained in Experiments 1 and 2. Consistent with earlier findings (Dannenbring & Bregman, 1978; Rasch, 1978), it is clear that asynchrony is a far better cue for segregating the components that are partly exposed (i.e., begin before or end after the others) than those that are not (i.e., begin after or end before the others). The relatively small contribution of harmonic relations between the Cs and Ds, which did not reach significance as a main effect, indicates that this factor is far less important than asynchrony in determining the fusion of complex tones of short duration. The above-chance performance observed for the nonoverlapping controls indicates that the mere presence of the Cs can be sufficient to identify the target rhythm. Presumably, either backward recognition masking weakens the representation of the Ds or the Cs act in a manner akin to a grace note to form a coherent DC unit. In either case, listeners would be able to select the regular targets (rhythm) from the irregular Ds based on differences in coloration or salience, even when the Ds and Cs are not fused per se. The role of such schema-based strategies in comparison with the effects of preattentive primitive grouping cues on RMR merits further investigation.

General Discussion

The Effect of Onset Asynchrony on Perceptual Organization Is Asymmetric

The results of Experiment 3 challenge an assumption that has remained prominent since the early auditory-grouping studies (e.g., Bregman & Pinker, 1978); namely that asynchrony, in particular that of onsets, is directly responsible for the observed segregation of temporally overlapping sounds into separate perceived events. When the degree of temporal overlap between the Ds and Cs was maintained constant at 48 ms (i.e., the total duration of the Ds), onset asynchrony affected fusion only weakly; indeed, for some listeners it had no effect at all. This suggests that onset asynchrony acts as a strong cue for the segregation of a brief target sound from a mixture only if the target is partly exposed by the asynchrony. Dannenbring and Bregman (1978) reported a similar asymmetry in the effect of onset asynchrony on the extent to which a component could be captured from a complex sound into a sequential stream. However, there may be another explanation for the apparent weakness of SOA as a grouping cue in Experiment 3.

The high level of performance in the RMR task for an SOA of 48 ms when the Ds and Cs overlap does not necessarily imply that they have fused perceptually. Indeed, the old-plus-new heuristic (Bregman, 1990) specifically assumes that a leading sound can be removed perceptually from a sound mixture to facilitate computation of the perceptual properties corresponding to the new acoustic elements. For example, preceding a flat spectrum with the spectral complement of a vowel (i.e., the inverse of the peaks and valleys in its normal spectrum) leads to the percept of a vowel-like coloration on the flat spectrum. This phenomenon is known as the enhancement effect (Summerfield, Haggard, Foster, & Gray, 1984). Although the operation of the old-plus-new heuristic improves the effective signal-to-noise ratio for the new sound, the separation is not a perfect one. In the context of Experiment 3, the perceptual properties of the Ds can be separated from the C plus D mixture on the basis of the independent sample of the leading Cs. However, the Ds are never heard in isolation and so are colored by their overlap, resulting in small but potentially discernible differences in properties like loudness and roughness compared with Ds heard in isolation. This offers an opportunity for schema-based selection strategies to uncover the target rhythm by exploiting the coloration of the Ds to tease them apart from the acoustically similar regular tones. To the extent that some listeners may have used a strategy of this kind, the effect of onset asynchrony on the fusion of temporally overlapping Cs and Ds may have been underestimated in Experiment 3.

Our findings illustrate the importance of attempting to tease apart the role of various temporal factors in simultaneous grouping, such as duration of overlap, onset synchrony, and offset synchrony. The studies of Bregman and Pinker (1978) and Turgeon et al. (2002) did not make such an attempt. However, the need to maintain temporal overlap constant as onset asynchrony was manipulated did not allow the teasing apart of the relative contributions of onset and offset asynchrony in Experiment 3. Although there is reason to expect a substantially smaller contribution of differences in offset than onset time (see, e.g., Roberts & Moore, 1991), a complete picture of the relative contribution of onset asynchrony, offset asynchrony, and duration of overlap to cross-spectral grouping would require two further experiments. One experiment should vary both temporal overlap and offset asynchrony while maintaining synchronous onsets, and the other should vary both temporal overlap and onset asynchrony while maintaining synchronous offsets.

Sound-Source Segregation Can Be Independent of Sound-Source Location in Space

Regardless of whether the tones came from a common location in space (free-field presentation) did not seem to matter in Experiment 1. In contrast, in Experiment 2 there was evidence for greater fusion of otherwise identical Cs and Ds when they were presented diotically rather than dichotically over headphones. Although it is conceivable to have real-world situations in which one sound stimulates one ear only, such situations are relatively rare. Free-field testing is more akin to real-world situations in which each of many individual sounds stimulate both ears, though at slightly different times and intensities, allowing for the computation of the location of each sound source. Within a free-field context, sound-source separation in space does not seem to be a potent cue for segregating brief sounds.

The present and prior RMR experiments (Turgeon et al., 2002) provide empirical support that spectrotemporal regularities are weighted more in auditory scene analysis than the spatial positions of sound sources. In particular, synchrony and temporal overlap appear to play the dominant role in grouping sounds. This is predicted from the theory of indispensable attributes (TIA), originally proposed by Kubovy (1981) and recently debated by Neuhoff (2003) and by Kubovy and his colleagues (Kubovy & Van Valkenburg, 2001; Van Valkenburg & Kubovy, 2003). According to TIA, auditory objects are formed in pitch time, whereas visual objects are formed in space time. The description of the retinal image in two spatial coordinates and of the auditory image in Frequency × Time coordinates (i.e., those of a spectrogram) might be best suited to describe different perceptual properties of world objects and events (what) and to specify different parameters of action (how). By this view, vision is specialized for spatiotemporal change and audition for spectrotemporal change. This is not to imply that their contributions are exclusive. Just as the acoustic signal provides some information to the animal about the displacement of a target sound source in space, the visual signal can provide information about spectrotemporal change, such as a change in color over time or a change in the flickering frequency of a light. Without taking a position regarding the specific tenets of TIA, our view, developed in Cisek and Turgeon (1999), shares with it the idea that the physical properties of visual and auditory signals motivate some degree of specialization.

Theoretical Considerations on Sound-Source Segregation (How Many) and Localization (Where)

The experiments reported here used Cs and Ds with spectrally interleaved components. However, the finding in a previous RMR study (Turgeon et al., 2002) that brief narrow-band noises presented at different spatial locations can fuse even when they are wide apart in frequency suggests that binaural fusion is not spectrally limited for the purpose of computing sound-source characteristics like pitch and timbre. This does not take into account whether the location of the source is perceived correctly; that is, in the present study, when a single sound was perceived, we do not know whether the source location in space (where) was perceived accurately. Further investigation is needed to establish whether the perception of a single sound event at a particular location (what is where) is spectrally limited. One way to proceed would be by asking separate questions for the same signal, namely, one evaluating fusion (e.g., How many sounds do you hear?) and another evaluating sound-source localization (e.g., Where are individual sounds coming from?). Some informal observations of Martine Turgeon, who is a highly trained listener with the RMR paradigm, suggest that there may be different processes for (a) retrieving the number and identity of sound sources, (b) retrieving the locations of sound sources, and that (c) our methodology is sensitive only to the former. When the irregular Ds fused clearly with simultaneous Cs on the other side of the semicircular array in Experiment 1, the rhythm was released from masking and perceived as coming from its veridical location. On the other hand, the irregular DC units were typically perceived as coming from an illusory location, namely, from a virtual source somewhere between the central speaker and the veridical location of the Cs, though somewhat closer to the latter. The perceived location of the virtual source of the DC complex was however more diffuse than that of the target rhythm (i.e., not as accurately defined in space).

The auditory system might localize sound sources in different ways. One is familiar and involves the computation of interaural time differences and interaural level differences for spectrally overlapping bands of energy across the two ears (Blauert, 1983; Jeffress, 1972; Lindeman, 1986). Alternatively, the system might attribute a single location to a perceived source when its spectrally distributed components exhibit properties that are typical of distinct world events, such as temporal correlation (e.g., synchronous onsets, common temporal envelope) and/or spectral regularities (e.g., harmonicity). One of the major conclusions of Experiment 1 is that when brief complex tones happen synchronously, perceptual fusion or segregation does not depend on sound-source separation in space. Sound segregation is such a basic property of audition that one might expect the system to compute it even in the face of ambiguity in the signal (e.g., as to where it comes from).

The Role of Attentionally Driven Grouping Strategies in RMR

The use of the RMR paradigm creates perceptual ambiguity in which the Ds can either group sequentially with the rhythm so as to form the masked-rhythm sequence or group simultaneously with the Cs, forming new DC units, perceptually distinct from the rhythmic sounds. This might be the ideal condition under which attentionally based strategies can jump in, resolving the ambiguity in the signal on the basis of cognitive expectations and/or learned heuristics. Indeed, the finding of above-chance performance for the nonoverlap controls in all of the experiments reported here supports this idea. This poses an important challenge for the use of the RMR paradigm to isolate primitive-grouping heuristics in auditory-unit formation, that is, biologically implemented hard-wired heuristics having evolved to extract adaptive regularities in the energy array. By the same token, such perceptual ambiguity makes RMR a good tool to investigate the interaction between primitive and schema-based grouping cues. Further RMR studies could pit preattentive and attentionally driven cues against each other.

Conclusions

The use of the RMR paradigm allows the evaluation of the relative importance of auditory-grouping cues for sound-source determination in the case of short signals. Together, the results of the three experiments support the conclusion that temporal coincidence and deviations from it, through increased asynchrony and/or decreased temporal overlap, play by far the most important role in the perceptual fusion of acoustic components of short duration. An onset-to-onset disparity between 30 and 40 ms appears to segregate partials of short duration that are partly overlapping in time, independent of their harmonic and spatial relations (Experiments 1 and 2). On the other hand, such onset-to-onset disparities do not appear to be sufficient for perceptual segregation when the Ds overlap with the Cs throughout their entire duration (Experiment 3). These results suggest that temporal overlap might have played a more important role in the reported perceptual fusion of past studies, such as that of the BC tones, attributed to their synchrony in the ABC paradigm of Bregman and Pinker (1978).

Consistent with earlier findings, it has been shown that the effect of simple harmonic ratios on the degree of fusion of brief tones is weak, regardless of whether they are asynchronous (Rasch, 1978) or synchronous (Moore, 1987). More surprisingly, it has also been established that harmonic relations can affect the fusion of brief nonoverlapping but temporally contiguous sounds. In terms of sound-source determination, evidence for the existence of an analysis of the dynamics of slow-intensity changes (temporal envelope) that crosses the spectrum and locations in space emerges from our present RMR results as well as from prior ones (Turgeon et al., 2002). Except under extreme conditions of spatial separation (dichotic presentation in Experiment 2), the fusion of brief tones appears to be independent of the spatial separation of their sources (Experiment 1).

To sum up, our finding that the degree of temporal overlap has a major effect on DC fusion probably reflects both the use of short-duration tones and the demands of the RMR task. The generality of this conclusion might be explored through the use of longer duration tones and strategies to reduce the potential usefulness of attentional cues.

References

Achim, A., Ahad, P. A., & Bregman, A. S. (1995). *Manager of Auditory Perception and Linguistic Experiments (MAPLE)*. Montreal, Quebec, Canada: McGill University, Department of Psychology, Auditory Perception Laboratory.

Blauert, J. (1983). *Spatial hearing.* Cambridge, MA: MIT Press.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA: MIT Press.

Bregman, A. S. (1993). Auditory scene analysis: Hearing in complex environments. In S. McAdams & E. Bigand (Eds.), *Thinking in sounds: Cognitive aspects of human audition* (pp. 10–36). Oxford, England: Oxford University Press.

Bregman, A. S., & Ahad, P. A. (1996). *Demonstrations of auditory scene analysis: The perceptual organization of sound* [CD and booklet] (pp. 41–42). Montreal, Quebec, Canada: McGill University, Department of Psychology, Auditory Perception Laboratory..

Bregman, A. S., & Pinker, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology, 32,* 19–31.

Broadbent, D. E., & Ladefoged, P. (1957). On the fusion of sounds reaching different sense organs. *Journal of the Acoustical Society of America, 29,* 708–710.

Brunstrom, J. M., & Roberts, B. (2000). Separate mechanisms govern the selection of spectral components for perceptual fusion and for the computation of global pitch. *Journal of the Acoustical Society of America, 107,* 1566–1577.

Carlyon, R. (2004). How the brain separates sounds. *Trends in Cognitive Sciences, 8,* 465–471.

Ciocca, V., & Darwin, C. J. (1999). The integration of nonsimultaneous frequency components into a single virtual pitch. *Journal of the Acoustical Society of America, 105,* 2421–2430.

Cisek, P., & Turgeon, M. (1999). Binding through the fovea: A tale of perception in the service of action. *Psyche, 5,* 5–34. Retrieved September 5, 2005, from http://psyche.csse.monash.edu.au/v5/psyche-5–34-cisek.html

Dannenbring, G., & Bregman, A. S. (1978). Streaming vs. fusion of sinusoidal components of complex tones. *Perception & Psychophysics, 24,* 369–376.

Darwin, C. J. (1997). Auditory grouping. *Trends in Cognitive Sciences, 1,* 327–333.

Darwin, C. J., & Carlyon, R. (1995). Auditory grouping. In B. C. J. Moore (Ed.), *Hearing: The handbook of perception and cognition* (Vol. 6, pp. 387–424). London: Academic Press.

Darwin, C. J., & Ciocca, V. (1992). Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component. *Journal of the Acoustical Society of America, 91,* 3381–3390.

Darwin, C. J., & Hukin, R. W. (1998). Perceptual segregation of a harmonic from a vowel by interaural time difference in conjunction with mistuning and onset asynchrony. *Journal of the Acoustical Society of America, 103,* 1080–1084.

Darwin, C. J., & Sutherland, N. S. (1984). Grouping frequency components of vowels: When is a harmonic not a harmonic? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 36*(A), 193–208.

de Boer, E. (1976). On the "residue" and auditory pitch perception. In W. D. Keidel & W. D. Neff (Eds.), *Handbook of sensory physiology* (Vol. 5, pp. 479–583). Berlin, Germany: Springer-Verlag.

Demany, L., & Semal, C. (1988). Harmonic and melodic octave templates. *Journal of the Acoustical Society of America, 88,* 2126–2135.

Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notch-noise data. *Hearing Research, 47,* 103–138.

Hartmann, W. M., McAdams, S., & Smith, B. K. (1990). Hearing a mistuned harmonic in an otherwise periodic complex tone. *Journal of the Acoustical Society of America, 88,* 1712–1724.

Henke, W. L. (1997). *MITSYN languages.* Belmont, MA: Author.

Hill, N. I., & Darwin, C. J. (1993). Lateralization of a perturbed harmonic: Effects of onset asynchrony and mistuning. *Journal of the Acoustical Society of America, 100,* 2352–2364.

Howell, D. A. (2002). *Statistical methods for psychology* (5th ed.). Duxbury, CA: Thomson Learning.

Hukin, R. W., & Darwin, C. J. (1995). Effects of contralateral presentation and of interaural time differences in segregating a harmonic from a vowel. *Journal of the Acoustical Society of America, 98,* 1380–1387.

Jeffress, L. A. (1972). Binaural signal detection: Vector theory. In J. V. Tobias (Ed.), *Foundations of modern auditory theory* (Vol. 2, pp. 349–370). New York: Academic Press.

Kidd, G., Mason, C., Rohtla, T. L., & Deliwala, P. S. (1998). Release from distracting due to spatial separation of sources in the identification of nonspeech auditory patterns. *Journal of the Acoustical Society of America, 104,* 422–431.

Kubovy, M. (1981). Concurrent-pitch segregation and the theory of indispensable attributes. In M. Kubovy & J. Pomerantz (Eds.), *Perceptual organization* (pp. 55–99). Hillsdale, NJ: Erlbaum.

Kubovy, M., & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition, 80,* 97–126.

Lindeman, W. (1986). Extension of a binaural cross-correlation model by contralateral inhibition: I. Simulation of lateralization for stationary signals. *Journal of the Acoustical Society of America, 80,* 1609–1622.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.* Cambridge, MA: MIT Press.

Massaro, D. W. (1975). Backward recognition masking. *Journal of the Acoustical Society of America, 58,* 1059–1065.

Moore, B. C. J. (1987). The perception of inharmonic complex tones. In W. A. Yost & C. S. Watson (Eds.), *Auditory processing of complex sounds* (pp. 180–189). Hillsdale, NJ: Erlbaum.

Moore, B. C. J., Glasberg, B. R., & Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *Journal of the Acoustical Society of America, 80,* 479–483.

Neuhoff, J. G. (2003). Pitch variation is unnecessary (and sometimes insufficient) for the formation of auditory objects. *Cognition, 87,* 219–224.

Patterson, R. D. (1973). The effects of relative phase and the number of components on residue pitch. *Journal of the Acoustical Society of America, 53,* 1565–1572.

Rasch, R. A. (1978). The perception of simultaneous notes such as in polyphonic music. *Acustica, 40,* 21–33.

Roberts, B., & Brunstrom, J. M. (1998). Perceptual segregation and pitch shifts of mistuned components in harmonic complexes and in regular inharmonic complexes. *Journal of the Acoustical Society of America, 104,* 2326–2338.

Roberts, B., & Brunstrom, J. M. (2001). Perceptual fusion and fragmentation of complex tones made inharmonic by applying different degrees of frequency shift and spectral stretch. *Journal of the Acoustical Society of America, 110,* 2479–2490.

Roberts, B., & Moore, B. C. (1991). The influence of extraneous sounds on the perceptual estimation of first-formant frequency in vowels under conditions of asynchrony. *Journal of the Acoustical Society of America, 89,* 2922–2932.

Summerfield, Q., Haggard, M., Foster, J., & Gray, S. (1984). Perceiving vowels from uniform spectra: Phonetic exploration of an auditory aftereffect. *Perception & Psychophysics, 35,* 203–213.

Turgeon, M., & 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. *Journal of the Acoustical Society of America, 102*(Suppl. 1), 3160.

Turgeon, M., Bregman, A. S., & Ahad, P. A. (2002). Rhythmic masking release: Contribution of cues for perceptual organization to the cross-spectral fusion of concurrent narrow-band noises. *Journal of the Acoustical Society of America, 111,* 1819–1831.

Van Valkenburg, D., & Kubovy, M. (2003). In defense of the theory of indispensable attributes. *Cognition, 87,* 225–233.

Weibull, W. A. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics, 18,* 292–297.

Yost, W. A. (1991). Auditory image perception and analysis: The basis for hearing. *Hearing Research, 56,* 8–18.

Yost, W. A. (1997). The cocktail party problem: Forty years later. In R. A. Gilkey & T. R. Anderson (Eds.), *Binaural and spatial hearing in real and virtual environments* (pp. 329–348). Hillsdale, NJ: Erlbaum.

Yost, W. A., Dye, R. H., Jr., & Sheft, S. (1996). A simulated "cocktail party" with up to three sound sources. *Perception & Psychophysics, 58,* 1026–1036.

Received April 27, 2000
Revision received August 11, 2004
Accepted March 15, 2005 ■