

Assessing the effects of temporal coherence on auditory stream formation through comodulation masking release

Simon Krogholt Christiansen^{a)}

Centre for Applied Hearing Research, Technical University of Denmark, DK-2800 Lyngby, Denmark

Andrew J. Oxenham

Departments of Psychology and Otolaryngology, University of Minnesota, Minneapolis, Minnesota 55455

(Received 13 September 2013; revised 29 March 2014; accepted 7 April 2014)

Recent studies of auditory streaming have suggested that repeated synchronous onsets and offsets over time, referred to as “temporal coherence,” provide a strong grouping cue between acoustic components, even when they are spectrally remote. This study uses a measure of auditory stream formation, based on comodulation masking release (CMR), to assess the conditions under which a loss of temporal coherence across frequency can lead to auditory stream segregation. The measure relies on the assumption that the CMR, produced by flanking bands remote from the masker and target frequency, only occurs if the masking and flanking bands form part of the same perceptual stream. The masking and flanking bands consisted of sequences of narrowband noise bursts, and the temporal coherence between the masking and flanking bursts was manipulated in two ways: (a) By introducing a fixed temporal offset between the flanking and masking bands that varied from zero to 60 ms and (b) by presenting the flanking and masking bursts at different temporal rates, so that the asynchronies varied from burst to burst. The results showed reduced CMR in all conditions where the flanking and masking bands were temporally incoherent, in line with expectations of the temporal coherence hypothesis. © 2014 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.4872300>]

PACS number(s): 43.66.Mk, 43.66.Dc, 43.66.Ba [EB]

Pages: 3520–3529

I. INTRODUCTION

An important task of the auditory system is to segregate different sound sources within natural acoustic environments. The ability to perceptually segregate competing sounds and selectively attend to individual sources over time has long been a topic of intense study (for reviews, see Bregman, 1990; Moore and Gockel, 2002; Carlyon and Gockel, 2008). Many experiments have relied on subjective evaluations of perceptual organization, for instance, by asking subjects how many “streams” they perceive. In recent years, an increased emphasis has been placed on more indirect, performance-based measures of auditory stream segregation (e.g., Micheyl and Oxenham, 2010). Measures of performance allow experimenters to eliminate, or at least control for, bias effects and also open up the possibility of studying perceptual organization in non-human species. The aim of the present study was to investigate the effects of temporal coherence on auditory stream formation using masking and masking release as indirect performance-based measures of perceptual organization.

Early studies suggested that “peripheral channeling,” or tonotopic separation produced initially by cochlear filtering, may provide the physiological underpinnings of the phenomenon known as “auditory streaming” (van Noorden, 1975; Hartmann and Johnson, 1991; Beauvois and Meddis, 1996; McCabe and Denham, 1997). According to this framework,

sounds that stimulate different populations of tonotopically tuned neurons are segregated into different streams, whereas sounds that stimulate the same neural population are integrated within a single perceptual stream (Fishman *et al.*, 2004; Micheyl *et al.*, 2005; Bee *et al.*, 2010). It has, however, been shown that dimensions other than tonotopic separation, such as fundamental-frequency (F0) differences (Vliegen and Oxenham, 1999; Vliegen *et al.*, 1999; Grimault *et al.*, 2000) or waveshape-induced timbre differences (Roberts *et al.*, 2002) can also induce streaming. Nonetheless, the principle of neural separation may still hold in populations of neurons that are sensitive to higher-level features, such as F0 or pitch (Bendor and Wang, 2005).

More recently, emphasis has been placed not only on the *spatial* separation of neural responses to sounds in a sequence, but also on the *temporal* relationships between them (e.g., Elhilali *et al.*, 2009; Shamma *et al.*, 2011; Micheyl *et al.*, 2013a; Micheyl *et al.*, 2013b). The finding that sounds repeatedly presented synchronously tend to form a single perceptual stream has been referred to as the principle of “temporal coherence” (e.g., Elhilali *et al.*, 2009). Although not explicitly accounted for in earlier neural models of streaming (e.g., Fishman *et al.*, 2004; Micheyl *et al.*, 2005), temporal coherence has been reported to be a relatively strong auditory grouping cue, which can bind together components even when they are relatively widely spaced in frequency.

In the study by Elhilali *et al.* (2009), the role of temporal coherence in grouping was assessed by measuring listeners’ ability to detect a small temporal asynchrony between two

^{a)}Author to whom correspondence should be addressed. Electronic mail: skch@elektro.dtu.dk

spectrally distant target tones that were preceded by a series of repeating tones at the same two frequencies as the target tones. Previous studies have shown that listeners are able to detect asynchronies of just a few milliseconds between spectral components of a complex tone that is perceived as a single auditory object when all components are synchronous (e.g., Zera and Green, 1993a,b, 1995), but that they cannot accurately judge the relative timing (Bregman and Campbell, 1971; Broadbent and Ladefoged, 1959; Neff *et al.*, 1982; Roberts *et al.*, 2002) or synchrony (Micheyl *et al.*, 2010) of sounds that fall into separate auditory streams. Thus Elhilali *et al.* (2009) used the thresholds from their asynchrony detection task as an indirect measure of perceptual grouping.

Elhilali *et al.* (2009) showed that when the pairs of preceding tones were presented synchronously (temporally coherent condition), the threshold for detecting asynchrony between the final two target tones was around 2–4 ms, whereas when the preceding tones were presented asynchronously (temporally incoherent condition), the threshold for detecting asynchrony was nearly an order of magnitude larger. This outcome is consistent with the idea that temporal coherence leads to perceptual grouping, even when the target tones are separated by a large frequency difference (15 semitones in this case).

In the same paper, Elhilali *et al.* (2009) reported physiological results obtained from the primary auditory cortex (AI) of awake but passive ferrets. The cortical units that responded to one or the other of the target tones did not show sensitivity to temporal coherence between the two tones, so that the human behavioral data could not be predicted from the ferret neural data without postulating an additional stage of neural processing that included the computation of temporal correlations of the activity from the AI units.

The discrepancy between the human behavioral and ferret neural data may be due to the presence of additional processing in non-primary cortical networks, as hypothesized by Elhilali *et al.* (2009). However, other alternatives exist. One possibility is that neural differences are observed at the level of AI only in situations where the subject is awake and attending to the stimuli; in the Elhilali *et al.* (2009) the ferrets were exposed passively and had no incentive to attend to the stimuli. A second possibility is that humans and ferrets perceive the stimuli differently, although behavioral studies to date suggest generally similar patterns of performance (Ma *et al.*, 2010). A third possibility is that the thresholds in the human behavioral task used by Elhilali *et al.* (2009) do not accurately reflect the perceptual organization of the stimuli. The task involved asynchrony detection in two conditions: In the temporally coherent condition, all the preceding tone pairs were synchronously gated, so the presence of the target resulted in the only stimulus asynchrony, whereas in the temporally incoherent condition, all the tone pairs were asynchronous, so that the target did not introduce the new “feature” of asynchrony to the stimulus. Thus it is possible that the behavioral results of Elhilali *et al.* (2009) were determined by the number of “distracting” asynchronies rather than by the perceptual organization of the target stimuli.

To test an alternative measure of streaming that does not suffer from the potential confounds associated with the asynchrony detection task used in Elhilali *et al.* (2009), we used comodulation masking release (CMR; Hall *et al.*, 1984). The term CMR refers to the finding that the detection of a target (usually a tone) in the presence of a masker with slow amplitude fluctuations can be improved when masker energy at remote frequencies has amplitude fluctuations that are coherent with those of the on-frequency masker. Different processing strategies have been proposed as underlying mechanisms for CMR. In general, two different categories exist: The first category involves within-channel processes, whereby changes in the amplitude envelope of a single masker band (or a broadband masker after filtering through a single auditory filter) can be used to explain CMR; the second category involves across-channel processes, where it is necessary to compare the amplitude envelopes at the outputs of multiple auditory filters to account for CMR. Many aspects of within-channel CMR can be explained to some extent by relatively peripheral auditory processes, such as suppression (e.g., Ernst and Verhey, 2008; Ernst *et al.*, 2010), or by modulation processing with a modulation filterbank following each auditory filter (Verhey *et al.*, 1999). In contrast, across-channel CMR requires processes such as across-channel envelope correlation (Richards, 1987), equalization-cancellation (Buus, 1985), dip-listening (Buus, 1985; Buss *et al.*, 2009), or an across-frequency comparison and integration of modulation information (Eddins and Wright, 1994; van de Par and Kohlrausch, 1998; Piechowiak *et al.*, 2007) to account for the data. An empirical feature that seems to distinguish within-channel from across-channel CMR is that across-channel CMR is affected by stimulus manipulations that are known to influence perceptual grouping (Grose and Hall, 1993; Dau *et al.*, 2005, 2009; Grose *et al.*, 2005; Grose *et al.* 2009; Verhey *et al.*, 2012). In particular, the coherent amplitude modulation in the remote-frequency (flanking) maskers seems only to aid signal detection in across-channel CMR when the flankers are thought to be perceptually grouped with the on-frequency masker. Other studies (Ernst and Verhey, 2008) have shown that CMR can also be observed with remote flanking bands (separated by as much as three octaves) in conditions that traditionally lead to a segregated percept; however, in the study by Ernst and Verhey (2008), the flanking bands were presented at a much higher intensity than the on-frequency band, and the observed CMR could have been the result of more peripheral effects, such as suppression, rather than across-channel processing (Ernst and Verhey, 2008; Ernst *et al.*, 2010).

In the present study, CMR was used as a measure of the perceptual organization of sounds under the assumption that the CMR produced by maskers remote in frequency from the target and on-frequency masker and with similar intensity will only occur if the masking and flanking bands form part of the same perceptual stream (e.g., Dau *et al.*, 2005, 2009). Perceptual streams were manipulated by embedding the masker and flankers that were synchronous with the target within a context of preceding and following maskers and flankers that were designed to lead to either perceptual

integration or segregation of the masking and the flanking noise bursts.

Through this measure, temporal coherence was investigated as a grouping cue, eliminating the potentially confounding influence of asynchronous stimuli within a task of asynchrony detection, such as that used by Elhilali *et al.* (2009). Temporal incoherence was introduced by manipulating the “gating envelope” in two ways, either through a constant asynchrony between the (on-frequency) masking and (off-frequency) flanking bursts or through different repetition rates for the masking and flanking bursts, leading to constantly varying asynchronies between the masking and flanking band onsets and offsets. In addition, the influence of the temporal coherence of the inherent fluctuations of the narrowband noises (“ongoing envelope”) was studied by using ongoing masker and flanker envelopes in the preceding and following bursts that were either correlated or uncorrelated.

II. EXPERIMENT 1: EFFECTS OF TEMPORAL INCOHERENCE AND GATING ASYNCHRONY ON CMR

A. Rationale

Across-channel CMR has been shown to depend on the perception of the masking and flanking bands within the same perceptual stream. Therefore the temporal coherence hypothesis predicts that CMR will be present when the maskers and flankers are preceded by coherent (i.e., synchronous) masker and flanker bursts and that CMR will be reduced or absent when the preceding masker and flanker bursts are presented incoherently or asynchronously, such that they form separate perceptual streams.

B. Method

1. Stimuli

Figure 1 shows a schematic representation of the stimuli used in experiment 1. The target signal (a 1-kHz pure tone) was embedded within a synchronously gated narrow-band (20-Hz-wide) masking noise (MN) centered at 1 kHz. Four flanking noises (FN) were presented synchronously with the target and MN, separated from the MN by ± 1 or 2 octaves (i.e., centered at 0.25, 0.5, 2, and 4 kHz). All FNs were also 20 Hz wide, and the ongoing envelope of each FN was either random or comodulated with that of the MN. Comodulation was achieved by generating the 20-Hz-wide Gaussian noise bands in the spectral domain and using the same amplitudes and phases at the different center frequencies. For the “random” configuration, each noise band was produced with independent randomly generated amplitudes and phases. The target and the noise bursts all had a duration of 187.5 ms, including 20 ms raised-cosine onset and offset ramps.

Prior to the presentation of the target tone and concurrent MN and FN bursts, a series of four precursors was presented (highlighted in light gray). These precursors consisted of noise bursts with the same average spectral and temporal properties as the FNs and MN. The number of precursors was chosen to correspond to the study by Dau *et al.* (2005). The time intervals between the onsets of successive noise

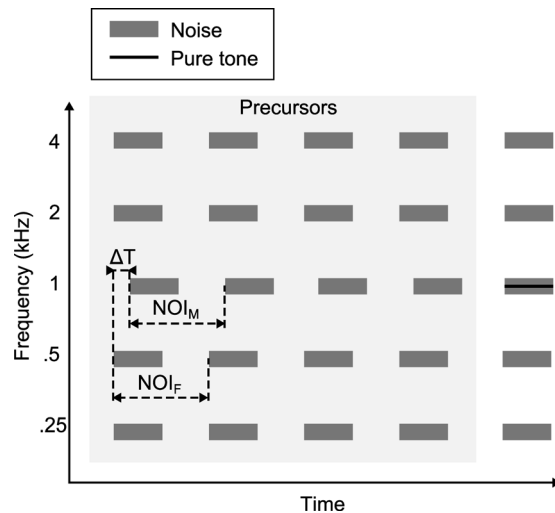


FIG. 1. Schematic representation of the stimuli used in experiment 1. Five narrow-band noises were presented at octave frequencies between 0.25 and 4 kHz. Each noise burst repeated five times. During the final noise burst a target signal (1-kHz pure tone) was embedded in the central noise band. The final noise bursts were always presented synchronously, but the temporal relationship between the central masking noise-band (MN) and the flanking noise-band (FN) precursors (highlighted in light gray) could be varied. The NOI_M and NOI_F represent the onset-to-onset time between successive MN and FN bursts, respectively. The value of ΔT represents the constant onset asynchrony between the MN and FN precursors in conditions where the NOI_M and NOI_F were equal. In the sketched example $NOI_M = NOI_F$ and $\Delta T \neq 0$ ms.

bursts (noise onset interval, NOI) were termed NOI_M and NOI_F for MN and FN bursts, respectively. All FNs were gated on and off simultaneously across frequency, but the timing of the MN precursors could vary independently from that of the FNs. Asynchronous gating of the MN, relative to the FNs, occurred in conditions where the MN precursors were delayed by ΔT relative to the FNs, or in conditions where $NOI_M \neq NOI_F$. Regardless of the temporal relationship between FN and MN precursors, the final FNs and MN, which were presented together with the target, were always gated on and off synchronously. For conditions with asynchronous precursors ($\Delta T \neq 0$), the NOI between the last precursor and the target interval deviated from the NOI between precursors to enable synchronized target noise bursts. This deviation was implemented by decreasing NOI_M by $\Delta T/2$ and increasing NOI_F by $\Delta T/2$ in the interval between the last precursor and the target. In conditions with either $\Delta T \neq 0$, or $NOI_M \neq NOI_F$, temporally overlapping portions of FN and MN had comodulated ongoing envelopes in the comodulated condition. This was realized by generating long-duration noises with comodulated ongoing envelopes and subsequently applying temporal windows to the noises to obtain the desired temporal gating properties of the precursors. The level of each narrow-band noise was set to 60 dB sound pressure level (SPL), and the level of the target was adaptively varied, as described in the following text, but was initially set to 75 dB SPL to ensure relatively easy detection of the target at the beginning of each adaptive track.

A total of seven different conditions were tested. In condition 1 (baseline), all precursors were synchronized and presented at a NOI_M and NOI_F of 250 ms with a ΔT of 0 ms.

Conditions 2-4 were again presented with NOI_M and NOI_F of 250 ms but with values of ΔT of 20, 40, and 60 ms, respectively. Conditions 5 and 6 kept the NOI_M constant at 250 ms, while the NOI_F was either 200 or 300 ms, respectively. Condition 7 had all precursors synchronized with the average NOI_M and NOI_F of 250 ms but with the NOI between each successive burst jittered by ± 30 ms with uniform distribution. Condition 7 thus had synchronized on- and offsets across frequency but not the temporal regularity of the baseline condition. All seven conditions were tested in both the random and comodulated configurations.

2. Procedure

An adaptive, three-interval, three-alternative forced-choice procedure was used together with a one-up two-down tracking rule to estimate the 70.7% correct point on the psychometric function (Levitt, 1971). The intervals were marked on a computer monitor, and feedback was provided after each trial. Listeners responded via computer keyboard or mouse. The initial step size of the target level was 8 dB, which was reduced to 4 and 2 dB after the second and fourth reversals, respectively. The adaptive run then continued for an additional six reversals at the final step size, and threshold was defined as the mean of the levels at those last six reversals. Four threshold estimates were obtained and averaged from each listener in each condition.

3. Listeners

Eight normal-hearing listeners, including the first author, participated in this experiment. The group consisted of four female and four male listeners, aged between 18 and 27 yr. The listeners (except the first author) were compensated monetarily for their participation at an hourly rate, and measurement sessions lasted 1–2 h including breaks. All listeners received 2–3 h of training in the same task before data collection began, and three to four sessions were required to complete the experiment. Data from one of the subjects were excluded from further analysis as the obtained thresholds did not stabilize through either initial training or through the data collection (intra-individual standard deviations remained around 7–10 dB). Thus the reported results are from the remaining seven subjects. The protocol was approved by the University of Minnesota's Institutional Review Board and the listeners provided written informed consent.

4. Apparatus

All stimuli were generated and presented through MATLAB (Mathworks, Natick, MA), using the AFC toolbox (Ewert, 2013). A sampling rate of 44.1 kHz was used, and the signals were presented through a personal computer with a 24-bit Lynx22 sound card (LynxStudio, Costa Mesa, CA). The stimuli were presented diotically through HD650 circumaural headphones (Sennheiser, Old Lyme, CT). The listeners were seated in a double-walled, sound-attenuating booth with a computer monitor that displayed instructions and feedback throughout the experiment.

C. Results and discussion

Threshold measurements were generally reliable within the seven subjects whose data were analyzed further; intra-individual standard deviations were typically between 0.5 and 2.5 dB and never exceeded 4 dB across the four estimates. In addition, the pattern of results was very similar across subjects, so only the mean data are reported here and are shown in Fig. 2. The top panels show the measured target thresholds in noise bands with random (squares) and comodulated (circles) ongoing envelopes, and the bottom panels show the amount of CMR, defined as the difference between thresholds in the random and comodulated configurations. The left column of Fig. 2 shows the results from conditions with a fixed NOI but varying degrees of asynchrony, ΔT ; the middle column shows the results from conditions with varying NOI_F ; and the right column shows the results from the condition with jittered NOIs. Results from the baseline condition ($\Delta T = 0$ ms, $\text{NOI}_M = \text{NOI}_F$) are shown in all three columns for comparison (hatched symbols and bars). The thresholds obtained in the baseline condition are consistent with results from Dau *et al.* (2005, 2009) using no precursor for both random and comodulated noise bands (hatched symbols). A one-way within-subjects (repeated-measures) analysis of variance (ANOVA) with threshold as the dependent variable and condition as the independent variable revealed no significant effect of precursor condition with the random-masker configuration (squares; upper panels) [$F(6,36) = 1.00$, $p = 0.44$]. In contrast, the conditions with the comodulated masker and flankers (circles; upper panels) showed a significant effect of precursor condition [$F(6,36) = 14.7$, $p < 0.001$].

The amount of CMR was treated as the dependent variable in another one-way repeated-measures ANOVA with condition as the factor. A main effect of precursor condition was found [$F(6,36) = 9.97$, $p < 0.001$]. *Post hoc* analyses of the CMRs within the three groupings illustrated in the three lower panels of Fig. 2 showed a significant reduction in CMR for ΔT of 40 and 60 ms, relative to the synchronized precursors (lower left panel), a significant reduction in CMR for conditions with $\text{NOI}_M \neq \text{NOI}_F$ relative to the $\text{NOI}_M = \text{NOI}_F$ (lower middle panel), and no significant effect of jittered NOIs (jitter) relative to the baseline condition (regular) (lower right panel).

The results indicate that increasing asynchrony, ΔT , leads to decreasing CMR, as would be expected if the asynchrony led to increased perceptual segregation between the MN and FNs. Previous studies have shown that onset/offset asynchronies larger than 20–40 ms lead to increased stream segregation (e.g., Turgeon *et al.*, 2002; Turgeon *et al.*, 2005; Bregman and Pinker, 1978; Micheyl *et al.*, 2013b; Christiansen *et al.*, 2014), in good agreement with the data from this study.

Similarly, the middle panels of Fig. 2 show that presenting the flanking precursors at a different rate from that of the masking precursors also leads to a significant reduction of CMR, again in line with predictions based on segregation based on temporal incoherence (Elhilali *et al.*, 2009). Last, the rightmost panels of Fig. 2 show that temporal irregularities in the form of jittered NOIs do not affect the amount of CMR relative to the baseline condition, indicating that the reduced CMR in conditions with onset/offset asynchronies

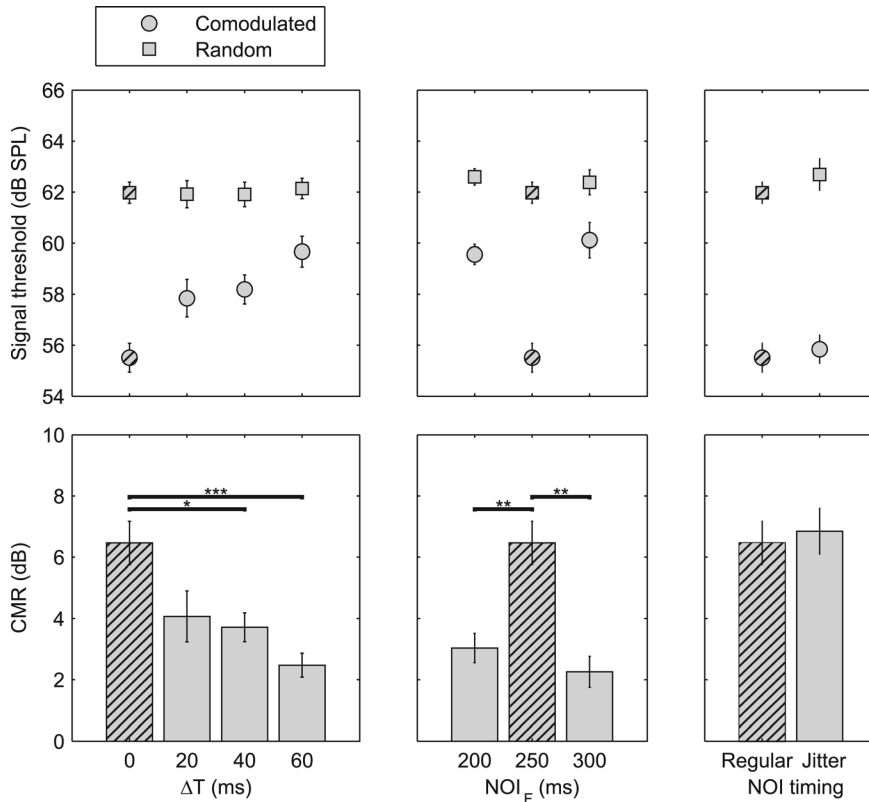


FIG. 2. Mean results from the seven subjects tested in experiment 1. The upper panels show the detection thresholds for the comodulated (circles) and random (squares) ongoing envelopes of the noise bursts. The bars in the lower panels show the amount of CMR, defined as the difference between the thresholds in the comodulated and random ongoing envelope conditions. The error bars represent ± 1 standard error of the mean across subjects. The baseline condition ($\Delta T = 0$ ms; $\text{NOI}_M = \text{NOI}_F = 250$ ms, no jitter) is indicated by hatching and is shown in all panels. The asterisks in the lower panels show the results of *post hoc* (Tukey's test) comparisons of CMR between conditions, where *, **, and *** indicate $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively. The leftmost panels show conditions with increasing ΔT but identical NOIs. The middle panels show the effect of varying NOI_F while keeping $\text{NOI}_M = 250$ ms. The rightmost panel show the effect of synchronized, but temporally irregular noise bursts.

or different MN and FN rates is not caused simply by the reduced temporal regularity of the stimuli.

For all seven precursor conditions, the threshold for the comodulated configuration was significantly lower than that for the random configuration [seven paired *t*-tests; $t(6) > 4.46$, $p < 0.003$ in all cases; significant after Bonferroni correction], indicating that none of the precursor conditions led to a complete elimination of CMR as might be expected if perceptual segregation of the masking and flanking bands was complete. This outcome differs from the results of Dau *et al.* (2005, 2009) and Verhey *et al.* (2012), who reported a complete elimination of CMR in conditions where the MN was perceptually segregated from the FN using repeated FN bursts as pre- or post-cursors. However, unlike the studies by Dau *et al.* (2005, 2009) and Verhey *et al.* (2012), experiment 1 had conflicting streaming cues as the temporal incoherence of the gating envelopes should facilitate a two-stream percept, while the coherent ongoing envelopes of the precursors during portions of temporal overlap may have promoted a one-stream percept (e.g., Hall and Grose, 1990). These conflicting streaming cues may have led to an incomplete perceptual segregation of FNs and MN, resulting in some residual CMR. Experiment 2 was designed to test this potential conflict between ongoing envelope cues and temporal onset and offset gating cues.

III. EXPERIMENT 2: INFLUENCE OF ONGOING ENVELOPE COMODULATION VERSUS GATING SYNCHRONY

A. Rationale

In experiment 1, the potential conflict of incoherent gating combined with coherent ongoing envelopes between the

masking and flanking bands may have resulted in incomplete perceptual segregation of the masker and flankers; this in turn may have resulted in residual CMR. To test this hypothesis, two of the precursor conditions from experiment 1 were retested, but with random ongoing envelopes on all (FN and MN) precursors, to eliminate any potential fusion due to ongoing comodulation within the precursors. The ongoing envelopes of the final MN and FNs (i.e., those presented simultaneously with the target tone) remained either comodulated or random, as in experiment 1. If comodulation of the ongoing envelopes within the precursors induced some perceptual fusion, then the use of random ongoing envelopes should result in a further reduction or elimination of CMR in conditions with asynchronously gated precursors.

B. Method

1. Stimuli and procedure

The stimuli were identical to those used in experiment 1 except that all precursors always had random ongoing envelopes, regardless of whether the temporal envelopes of the MN and FN presented simultaneously with the target signal were comodulated or random. Only two precursor configurations were tested: $\Delta T = 0$ (baseline condition) and $\Delta T = 60$ ms (maximum onset/offset asynchrony from experiment 1). In both conditions, the NOI_M and NOI_F were 250 ms. The procedure and equipment were identical to those of experiment 1.

2. Listeners

Ten normal-hearing listeners participated in this experiment. Five of the listeners had also participated in

experiment 1 (including the first author). The group consisted of five female and five male listeners, aged between 19 and 34 yr. The listeners were compensated monetarily for their participation at an hourly rate, and measurement sessions lasted between 1 and 2 h including breaks. All listeners received at least 1 h of training in the same task before data collection began, and one to two sessions were required to complete the experiment.

C. Results and discussion

The results from experiment 2 are shown in Fig. 3 together with the results from experiment 1 for the corresponding precursor configurations. As in experiment 1, the intra-individual standard deviations were relatively small (0.5–2 dB, rarely exceeding 4 dB), and all subjects showed similar patterns of results, so only the mean data are shown here. The left panel shows the measured target detection thresholds for random (squares) and comodulated (circles) noises, and the right panel shows the CMR (the difference between random and comodulated thresholds). In both panels, the gray and open symbols indicate results from experiments 1 and 2, respectively. Note that there is some, but not complete, overlap between the subjects in the two experiments.

Mixed-model ANOVAs were carried out separately for the random and comodulated configurations, with threshold as the dependent variable, experiment as a between-subjects factor, and ΔT as a within-subjects factor.¹ The results of the ANOVA for the random configuration showed no significant effect of experiment (1 or 2) [$F(1,15) = 0.39$, $p = 0.54$] or

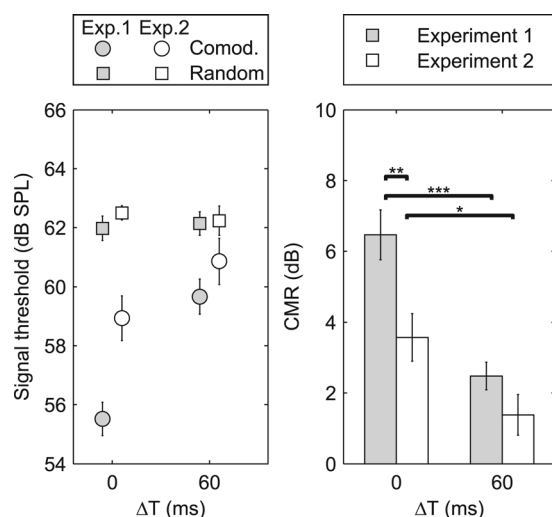


FIG. 3. Results from experiment 2 (open symbols), with data from experiment 1 replotted for comparison (gray symbols) for conditions with ΔT of 0 and 60 ms. The left panel shows the detection thresholds for the comodulated (circles) and random (squares) ongoing envelopes of the final noise bursts. The bars in the right panel show the CMR, and the error bars indicate ± 1 standard error of the mean between subjects. The asterisks in the right panel show the results of *post hoc* (Tukey-Kramer) comparisons of CMR, where *, **, and *** indicate $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively. In experiment 1, both the precursors and the noise bursts concurrent with the target signal had either random or comodulated ongoing envelopes. In experiment 2, the precursors always had random ongoing envelopes, and only the noise bursts concurrent with target signal were random or comodulated.

ΔT [$F(1,15) = 0.01$, $p = 0.91$] and no interaction [$F(1,15) = 0.36$, $p = 0.55$]. The absence of an effect of experiment was expected as the stimulus properties were identical across the two experiments. For the comodulated configuration, significant main effects were found for both experiment [$F(1,15) = 6.41$, $p = 0.02$] and ΔT [$F(1,15) = 34.95$, $p < 0.001$] along with a significant interaction [$F(1,15) = 4.70$, $p = 0.047$]. *Post hoc* *t*-tests indicated that the threshold increased significantly in experiment 2 relative to experiment 1 for $\Delta T = 0$ ms [*t*-test; $t(15) = 3.52$, $p < 0.01$], but the small increase for $\Delta T = 60$ ms was not significant [$t(15) = 1.19$, $p = 0.14$].

A similar mixed-model ANOVA of the CMRs revealed a significant effect of ΔT [$F(1,15) = 26.21$, $p < 0.001$], a significant effect of experiment [$F(1,15) = 9.363$, $p < 0.01$], but no significant interaction effect [$F(1,15) = 2.215$, $p = 0.157$]. The results of *post hoc* comparisons are indicated in the right panel of Fig. 3. The pair-wise comparisons showed first that regardless of whether the ongoing envelopes of the precursors were comodulated or not (experiment 1 vs experiment 2), the onset asynchrony $\Delta T = 60$ ms significantly reduced the amount of CMR. Second, the reduction in CMR between experiments 1 and 2 was significant for $\Delta T = 0$ ms but not for $\Delta T = 60$ ms. Under the assumption that the amount of CMR reflects the strength of perceptual fusion between the masking and flanking bands, this result indicates that random ongoing envelope fluctuations of the precursors reduce the fusion between MN and FNs, even though they have synchronized on- and offsets. The difference in CMR between $\Delta T = 0$ ms and $\Delta T = 60$ ms in experiment 2 shows that onset synchrony still provides a strong grouping cue when the precursors are not comodulated. Even though the comodulation of the ongoing envelopes of the precursors was removed in experiment 2, the signal thresholds were still significantly lower for the comodulated configuration than for the random configuration for both $\Delta T = 0$ ms [paired *t*-test; $t(9) = 5.30$, $p < 0.001$] and $\Delta T = 60$ ms [paired *t*-test; $t(9) = 2.39$, $p = 0.02$], suggesting that the fusion of MN and FNs was not eliminated. The results of experiment 2 thereby support the hypothesis that the comodulation of the precursors in experiment 1 provided a grouping cue that limited the streaming effects of the incoherently gated precursors. However, the results also show that the comodulation of the ongoing envelope of the precursors cannot explain why the CMR effect persisted for precursor conditions that were predicted to lead to a segregated percept.

IV. EXPERIMENT 3: EFFECT ON STREAMING OF EMBEDDING THE TARGET BETWEEN PRE- AND POST-CURSORS

A. Rationale

Experiment 1 showed that the presence of asynchronously gated precursors reduced the amount of CMR but did not fully eliminate it. Experiment 2 investigated the contribution of comodulation between the ongoing envelopes of overlapping portions of the precursor flanking and masking bands. Although precursors with random temporal envelopes produced less CMR than precursors with comodulated

envelopes when gated synchronously, even the precursors with random ongoing envelopes did not completely eliminate CMR when they were gated asynchronously.

In all the experimental conditions tested so far, the target was always presented in the final masker burst. It may be that listeners were able to develop a strategy of “ignoring” the precursors and focusing primarily on the final noise burst. It is known that switching attention can lead to breakdown of auditory stream segregation and lead to more fused percepts in alternating tone sequences (e.g., Carlyon *et al.*, 2001; Cusack *et al.*, 2004). Therefore switching attention to the sounds directly prior to the final noise burst might be an advantageous strategy, as it would lessen the possibility that the masking and flanking bands form separate perceptual streams. In this final experiment, the target was presented in an unpredictable location within a longer series of noise bursts. Because of that listeners were no longer able to ignore the precursors and had to monitor the repeated bursts to detect the presence of the signal. The hypothesis was that forcing listeners to attend to the longer sequence would lead to greater buildup of stream segregation and may thus lead to the elimination of CMR in cases where the flanking and masking bands were gated on and off incoherently.

B. Method

1. Stimuli and procedure

Figure 4 shows a schematic representation of the stimuli presented in each interval of a trial. The general structure of the stimuli was the same as that used in experiment 1 except that instead of 5 repeating noise bursts there were now 10 repetitions. In addition, the target was no longer embedded in the final noise burst but was instead presented synchronously with the fifth, sixth, or seventh noise burst, selected at

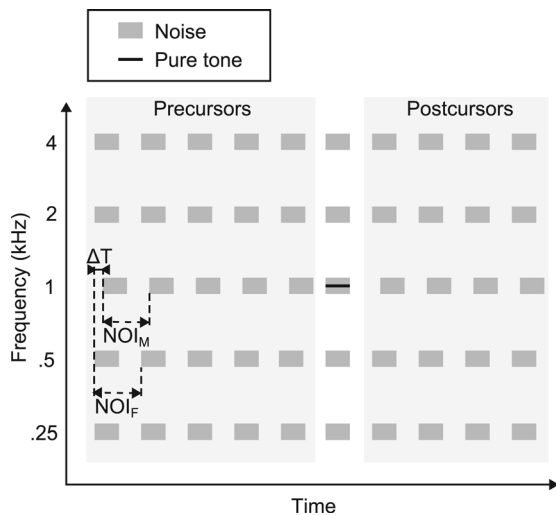


FIG. 4. Schematic representation of the stimuli used in experiment 3. Five narrow-band noise bursts were presented at octave frequencies between 0.25 and 4 kHz. Each noise burst was repeated 10 times, and the 1-kHz target was presented simultaneously with the 5th, 6th, or 7th noise burst, chosen at random in each trial. The noise bursts presented together with the target were always presented synchronously, but the onset asynchrony (ΔT) between the masking noise band and the flanking noise bands' pre- and post-cursors was either 0 or 60 ms. In the sketched example, the target signal is presented in the 6th interval.

random on each trial. The MN and FNs were always presented synchronously during the noise burst containing the target, and the non-target intervals in each trial had the MN and FNs synchronized on the same noise burst (fifth, sixth, or seventh) as in the target interval within a given trial. The MN and FN preceding and following the target (pre- and post-cursors, highlighted in light gray) were either synchronized ($\Delta T = 0$ ms) or had an asynchrony of $\Delta T = 60$ ms. In both conditions, the NOI_M and NOI_F were both set to 250 ms. The procedure and the set up were identical to that used in experiment 1.

2. Listeners

Eight normal-hearing listeners participated in this experiment. Five of the listeners had also participated in both experiments 1 and 2 (including the first author). The group consisted of four female and four male listeners, aged between 19 and 31 yr. The listeners were compensated monetarily for their participation at an hourly rate, and measurement sessions lasted between 1 and 2 h, including breaks. All listeners received at least 1 h of training in the same task before data collection began, and one to two sessions were required to complete the experiment.

C. Results

The individual results showed within-subject standard deviations that were typically around 0.5–2 dB and never exceeded 4 dB. In addition, all subjects showed a similar pattern of results across the different conditions, and so only the mean data are reported here. The mean data are shown in Fig. 5 together with the results replotted from experiment 1

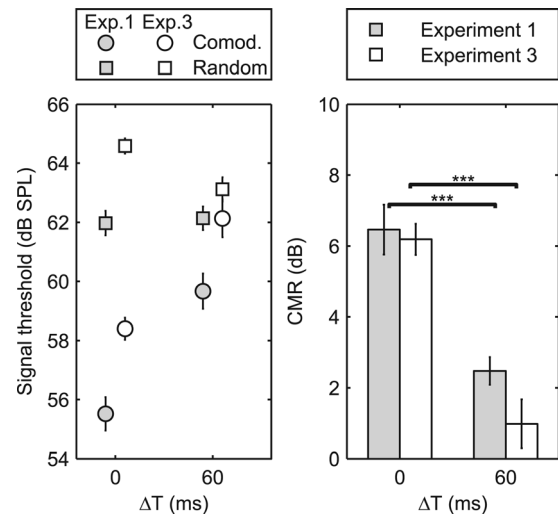


FIG. 5. Results from experiment 3 (open symbols), with data from experiment 1 replotted for comparison (gray symbols) for conditions with ΔT of 0 and 60 ms. The left panel shows the detection thresholds for the comodulated (circles) and random (squares) ongoing envelopes of the noise bursts. The bars in the right panel show the CMR, and the error bars indicate ± 1 standard error of the mean between subjects. The asterisks in the right panel show the results of *post hoc* (Tukey-Kramer) comparisons of CMR, where *** indicates $p < 0.001$. In experiment 1, the stimuli only contain pre-cursor noise bursts and target interval as depicted in Fig. 1. In experiment 3 the stimuli contain both pre- and post-cursors before and after the target interval, as depicted in Fig. 4.

for the corresponding precursor configurations. The left panel shows the detection thresholds from the random (squares) and comodulated (circles) configurations, and the right panel shows the CMR (difference between the random and comodulated thresholds). In both panels, the gray and open symbols indicate results from experiments 1 and 3, respectively.

A mixed-model ANOVA on thresholds in the random configuration showed a significant effect of experiment [$F(1,13) = 14.57, p < 0.01$], indicating that the addition of post-cursors and/or the randomized location of the target affected performance in experiment 3, relative to that in experiment 1. The analysis also showed a significant effect of ΔT [$F(1,13) = 7.59, p < 0.02$] and interaction (experiment by ΔT) [$F(1,13) = 11.97, p < 0.01$]. *Post hoc* analyses (paired *t*-tests) showed that detection thresholds with the random maskers were significantly poorer in the synchronized condition ($\Delta T = 0$ ms) than in the asynchronous condition ($\Delta T = 60$ ms) in experiment 3 [$t(7) = 3.28, p < 0.01$]. A mixed-model ANOVA on thresholds in the comodulated configuration revealed a significant effect of experiment [$F(1,13) = 23.47, p < 0.001$] and ΔT [$F(1,13) = 51.50, p < 0.001$] but no interaction [$F(1,13) = 0.14, p = 0.71$], indicating that the addition of post-cursors and/or the randomizing of the target location resulted in a similar increase in signal threshold for both the synchronous and asynchronous conditions.

The elevated thresholds observed in experiment 3 relative to experiment 1 may be due to an increased signal uncertainty in time. Green and Weber (1980) and Bonino and Leibold (2008) investigated the effect of temporal uncertainty in a detection task involving a 1 kHz pure tone target in a noise masker. Both studies found increased detection thresholds of 2–3 dB when going from a temporally certain to an uncertain position of the target; this is consistent with the observed increase in detection threshold for $\Delta T = 0$ ms for both the comodulated and random configurations, and for the $\Delta T = 60$ ms in the comodulated configuration. However, for the $\Delta T = 60$ ms in the random configuration, the increase is only about 1 dB. One possible explanation for the lower threshold for $\Delta T = 60$ ms relative to $\Delta T = 0$ ms in the random configuration may be that the synchrony of the noise bursts at the target location helped listeners identify where the target is likely to appear, effectively reducing the temporal uncertainty. If the temporal uncertainty is reduced for the $\Delta T = 60$ ms, a similar benefit would be expected for the comodulated configuration. The 3 dB increase for the $\Delta T = 60$ ms comodulated condition may therefore be a combination of a reduced ability to use across-channel information due to a perceptual segregation as well as an increased temporal uncertainty.

A further mixed-model ANOVA performed on the CMR values revealed a significant effect of ΔT [$F(1,13) = 70.6, p < 0.001$] but no significant effect of experiment [$F(1,13) = 2.14, p = 0.17$] or interaction (experiment by ΔT) [$F(1,13) = 1.24, p = 0.29$], indicating that the addition of post-cursors did not significantly affect the amount of CMR.

The original hypothesis was that stream segregation might be increased (and CMR reduced) by embedding the

target within a longer stream of noise bursts when the MN and FN bursts were asynchronous. The lack of an effect of (or interaction with) experiment on CMR is not consistent with this hypothesis. On the other hand, when considering just the results from experiment 3, the amount of CMR was significantly greater than zero for the synchronous condition ($\Delta T = 0$ ms) [paired *t*-test; $t(7) = 14.1, p < 0.001$], whereas no significant CMR was found in the asynchronous condition ($\Delta T = 60$ ms) [paired *t*-test; $t(7) = 1.42, p = 0.10$]. Thus in contrast to experiments 1 and 2, and consistent with the original hypothesis, no significant CMR was observed in the condition where the pre- and post-cursors were not temporally coherent, suggesting that the MN and FN bursts were sufficiently segregated to eliminate measurable CMR.

V. GENERAL DISCUSSION

A. Summary of results

Detection of a 1-kHz tone was measured in narrow bands of noise that were spaced at octave frequencies from 250 Hz to 4 kHz. The bands of noise were either comodulated (shared the same ongoing envelope) or independently generated. The noise bursts that were gated synchronously with the target tone were preceded or temporally surrounded by a series of noise bursts, intended to influence the perceptual organization of the sequence. The first experiment showed that temporally coherent on-frequency masker and off-frequency flanker noise bursts produced CMR and that reducing the coherence through a fixed asynchrony or through different presentation rates in the preceding noise bursts led to a reduction of CMR. The second experiment explored the effect of comodulation within the precursor bursts and found that random (independent) ongoing envelopes in the precursors reduced CMR when the precursors were synchronously gated, suggesting that incoherence in the ongoing portions of the precursor temporal envelopes can reduce fusion, even when the bursts are synchronously gated. The third experiment added “postcursors” that followed the target burst, in addition to the precursors, and randomized the position of the target, so that subjects were obliged to attend to more of the sequence. In general, thresholds were somewhat higher in experiment 3, and the amount of CMR was not significantly greater than zero when the precursors were asynchronously gated, suggesting that the MN and FN bursts were perceptually segregated from each other.

The results of all three experiments support the hypothesis that temporal coherence between noise bursts widely separated in frequency leads to the formation of a single perceptual stream, as evidenced by the finding of significant CMR in conditions where the masking and flanking bands were presented synchronously across bursts. Also consistent with the hypothesis was the finding that CMR was reduced or absent in conditions where the on-frequency and flanking pre- and post-cursors were not temporally coherent, either through a fixed asynchrony or through different presentation rates. The outcomes do not depend on the temporal regularity of the precursors, as temporally jittered (but synchronously gated) precursors produced as much CMR as the regular sequence of precursors.

Overall, the results provide support for the hypothesis of Elhilali *et al.* (2009) that temporal coherence plays an important role in the auditory streaming of widely separated frequency components, using a paradigm (CMR) that does not suffer from the potential confound of the asynchrony-detection task used by Elhilali *et al.* (2009). In addition, the results provide further support for the idea that across-channel CMR provides a viable indirect measure for investigating the perceptual organization of sounds.

B. Relation to previous studies and interpretations of perceptual segregation

Grose *et al.* (2005) measured CMR with maskers and flankers that were comodulated for the duration of the 400-ms target but otherwise had random temporal envelopes. They found that the introduction of these random temporal “fringes” significantly reduced CMR even though the maskers and flankers that were simultaneously present with the target tone were unchanged. They argued that the ongoing random noise may put “the system in a state where comodulation is not expected, and therefore potential cueing mechanisms (perhaps based on grouping by common modulation) are not activated.” Our observation in experiment 2 that random ongoing envelopes in the flankers, even when gated synchronously with the precursor masker bands, led to reduced CMR is consistent with the findings and conclusions of Grose *et al.* (2005). In fact, it is possible to consider the inherent fluctuations of the 20-Hz-wide noise bands and their gating on and off as coherent or incoherent modulation in two modulation-frequency regions: The gating involved a period of 250 ms (or 4 Hz), whereas 20-Hz-wide noise bands have modulation energy out to 20 Hz. Thus both the ongoing (inherent) modulation and the gating can be considered cases of temporal coherence. Within this framework, the temporal coherence at the level of the gating modulation, as well as the temporal coherence at the level of the inherent noise fluctuations, influences the perceptual organization of the sound. Thus both our results and those of Grose *et al.* (2005) can be understood as special cases of the general principle that streams fuse if they are temporally coherent but tend to segregate if they are incoherent at one or more levels of modulation analysis.

A somewhat surprising finding was that it was not possible to completely eliminate the CMR in most of the experimental conditions, given that Dau *et al.* (2005, 2009) and Verhey *et al.* (2012) found no across-channel CMR in conditions with pre- or post-cursors. One major difference between the current study and the studies by Dau *et al.* and Verhey *et al.* was that the earlier studies only presented FN pre- or post-cursors and no MN pre- or post-cursors. Therefore their stimuli did not have any ambiguous streaming cues, such as overlapping (and sometimes comodulated) portions of the pre- and post-cursor noise bursts. Removing comodulation within the precursors in experiment 2 decreased the CMR but did not eliminate it completely. Even in experiment 3, where the target was embedded between pre- and post-cursors and its position randomized, although the remaining CMR was not significantly different

from zero, it was also not significantly different from that found in experiment 1, leaving the result somewhat ambiguous.

It may be that even precursors with random ongoing envelopes, and onset asynchronies of up to 60 ms, are not enough to completely eliminate perceptual fusion and that the remaining temporal overlap of the MN and FN precursors (for 68% of their duration) acted as a grouping cue. Another possibility is that the stimuli were too short allow a sufficient buildup of stream segregation. Although several studies (e.g., Anstis and Saida, 1985; Bee *et al.*, 2010) suggest that the “buildup” of auditory streams takes place on a timescale of several seconds, the study of Dau *et al.* (2005) showed a complete elimination of CMR using a relatively short build-up period of only four precursors (for a total duration of 1 s). According to Moore and Gockel (2002), the extent to which sequential stream segregation occurs is directly related to the degree of perceptual difference between successive sounds: In the original study by Dau *et al.* (2005), the perceptual difference between MN and FNs was likely larger due to the absence of MN precursors. This may have led to faster stream segregation in Dau *et al.* (2005) than in the present study, and it remains possible that a complete elimination of CMR would have been observed in the present study if more precursors had been used.

ACKNOWLEDGMENTS

This work was supported by the Oticon Foundation and by NIH Grant No. R01 DC007657.

¹Although some subjects participated in both experiments, they were treated as independent for the purposes of this analysis to avoid problems of missing values. Treating the subjects as independent across experiment likely results in a loss of statistical power, making the current analysis a relatively conservative test of significance.

- Anstis, S., and Saida, S. (1985). “Adaptation to auditory streaming of frequency-modulated tones,” *J. Exp. Psychol. Hum. Percept. Perform.* **11**, 257–271.
- Beauvois, M. W., and Meddis, R. (1996). “Computer simulation of auditory stream segregation in alternating tone sequences,” *J. Acoust. Soc. Am.* **99**, 2270–2280.
- Bee, M. A., Micheyl, C., Oxenham, A. J., and Klump, G. M. (2010). “Neural adaptation to tone sequences in the songbird forebrain: Patterns, determinants, and relation to the build-up of auditory streaming,” *J. Comp. Physiol. A* **196**, 543–557.
- Bendor, D., and Wang, X. (2005). “The neuronal representation of pitch in primate auditory cortex,” *Nature* **436**, 1161–1165.
- Bonino, A., and Leibold, L. J. (2008). “The effect of signal-temporal uncertainty on detection in bursts of noise or a random-frequency complex,” *J. Acoust. Soc. Am.* **124**, EL321–EL327.
- Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA), pp. 1–736.
- Bregman, A. S., and Campbell, J. (1971). “Primary auditory stream segregation and perception of order in rapid sequences of tones,” *J. Exp. Psychol.* **89**, 244–249.
- Bregman, A. S., and Pinker, S. (1978). “Auditory streaming and the building of timbre,” *Can. J. Psychol.* **32**, 19–31.
- Broadbent, D. E., and Ladefoged, P. (1959). “Auditory perception of temporal order,” *J. Acoust. Soc. Am.* **31**, 1539–1540.
- Buss, E., Grose, J. H., and Hall, J. W. (2009). “Features of across-frequency envelope coherence critical for comodulation masking release,” *J. Acoust. Soc. Am.* **126**, 2455–2466.
- Buus, S. (1985). “Release of masking caused by envelope fluctuation,” *J. Acoust. Soc. Am.* **78**, 1958–1965.

- Carlyon, R. P., Cusack, R., and Foxton, J. M. (2001). "Effects of attention and unilateral neglect on auditory stream segregation," *J. Exp. Psychol. Hum. Percept. Perform.* **27**, 115–127.
- Carlyon, R. P., and Gockel, H. E. (2008). "Effects of harmonicity and regularity on the perception of sound sources" in *Auditory Perception of Sound Sources*, edited by W. A. Yost (Springer, New York), Chap. 7, pp. 191–213.
- Christiansen, S. K., Jepsen, M. L., and Dau, T. (2014). "Effects of tonotopicity, adaptation, modulation tuning, and temporal coherence in 'primitive' auditory stream segregation," *J. Acoust. Soc. Am.* **135**, 323–333.
- Cusack, R., Deeks, J., Aikman, G., and Carlyon, R. P. (2004). "Effects of location, frequency region, and time course of selective attention on auditory scene analysis," *J. Exp. Psychol. Hum. Percept. Perform.* **30**, 643–656.
- Dau, T., Ewert, S., and Oxenham, A. J. (2005). "Effects of concurrent and sequential streaming in comodulation masking release," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigne, S. McAdams, and L. Collet (Springer-Verlag, Berlin).
- Dau, T., Ewert, S., and Oxenham, A. J. (2009). "Auditory stream formation affects comodulation masking release retroactively," *J. Acoust. Soc. Am.* **125**, 2182–2188.
- Eddins, D. A., and Wright, B. A. (1994). "Comodulation masking release for single and multiple rates of envelope fluctuation," *J. Acoust. Soc. Am.* **96**, 3432–3442.
- Elhilali, M., Ling, C., Micheyl, C., Oxenham, A. J., and Shamma, S. A. (2009). "Temporal coherence in the perceptual organization and cortical representation of auditory scenes," *Neuron* **61**, 317–329.
- Ernst, S. M. A., Rennie, J., Kollmeier, B., and Verhey, J. L. (2010). "Suppression and comodulation masking release in normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **128**, 300–309.
- Ernst, S. M. A., and Verhey, J. L. (2008). "Peripheral and central aspects of auditory across-frequency processing," *Brain Res.* **1220**, 246–255.
- Ewert, S. D. (2013). "AFC—A modular framework for running psychoacoustic experiments and computational perception models," in *Proceedings of the International Conference on Acoustics AIA-DAGA2013*, Merano, Italy, pp. 1326–1329.
- Fishman, Y., Arezzo, J. C., and Steinschneider, M. (2004). "Auditory stream segregation in monkey auditory cortex: Effects of frequency separation, presentation rate, and tone duration," *J. Acoust. Soc. Am.* **116**, 1656–1670.
- Green, D. M., and Weber, D. L. (1980). "Detection of temporally uncertain signals," *J. Acoust. Soc. Am.* **67**, 1304–1311.
- Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., and Collet, L. (2000). "Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency," *J. Acoust. Soc. Am.* **108**, 263–271.
- Grose, J. H., Buss, E., and Hall, J. W. (2009). "Within- and across-channel factors in the multiband comodulation masking release paradigm," *J. Acoust. Soc. Am.* **125**, 282–293.
- Grose, J. H., and Hall, J. W. (1993). "Comodulation masking release: Is comodulation sufficient?," *J. Acoust. Soc. Am.* **93**, 2896–2902.
- Grose, J. H., Hall, J. W., Buss, E., and Hatch, D. R. (2005). "Detection of spectrally complex signals in comodulated maskers: Effect of temporal fringe," *J. Acoust. Soc. Am.* **118**, 3774–3782.
- Hall, J. W., and Grose, J. H. (1990). "Comodulation masking release and auditory grouping," *J. Acoust. Soc. Am.* **88**, 119–125.
- Hall, J. W., Haggard, M. P., and Fernandes, M. A. (1984). "Detection in noise by spectro-temporal pattern analysis," *J. Acoust. Soc. Am.* **76**, 50–56.
- Hartmann, W. M., and Johnson, D. (1991). "Stream segregation and peripheral channeling," *Music Percept.* **9**, 155–184.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Ma, L., Micheyl, C., Yin, P., Oxenham, A. J., and Shamma, S. (2010). "Behavioral measures of auditory streaming in ferrets (*Mustela putorius*)," *J. Comp. Psychol.* **124**, 317–330.
- McCabe, S., and Denham, M. J. (1997). "A model of auditory streaming," *J. Acoust. Soc. Am.* **101**, 1611–1621.
- Micheyl, C., Hanson, C., Demany, L., Shamma, S., and Oxenham, A. J. (2013a). "Auditory stream segregation for alternating and synchronous tones," *J. Exp. Psychol. Hum. Percept. Perform.* **39**, 1568–1580.
- Micheyl, C., Hunter, C., and Oxenham, A. J. (2010). "Auditory stream segregation and the perception of across-frequency synchrony," *J. Exp. Psychol. Hum. Percept. Perform.* **36**, 1029–1039.
- Micheyl, C., Kreft, H., Shamma, S., and Oxenham, A. J. (2013b). "Temporal coherence versus harmonicity in auditory stream formation," *J. Acoust. Soc. Am.* **133**, EL188.
- Micheyl, C., and Oxenham, A. J. (2010). "Objective and subjective psychophysical measures of auditory stream integration and segregation," *J. Assoc. Res. Otolaryngol.* **11**, 709–724.
- Micheyl, C., Tian, B., Carlyon, R. P., and Raunschecker, J. P. (2005). "Perceptual organization of tone sequences in the auditory cortex of awake macaques," *Neuron* **48**, 139–148.
- Moore, B. C. J., and Gockel, H. E. (2002). "Factors influencing sequential stream segregation," *Acta. Acust. Acust.* **88**, 320–332.
- Neff, D. L., Jesteadt, W., and Brown, E. L. (1982). "The relation between gap discrimination and auditory stream segregation," *Percept. Psychophys.* **31**, 493–501.
- Piechowiak, T., Ewert, S. D., and Dau, T. (2007). "Modeling comodulation masking release using an equalization-cancellation mechanism," *J. Acoust. Soc. Am.* **121**, 2111–2126.
- Richards, V. M. (1987). "Monaural envelope correlation perception," *J. Acoust. Soc. Am.* **82**, 1621–1630.
- Roberts, B., Glasberg, B. R., and Moore, B. C. J. (2002). "Primitive stream segregation of tone sequences without differences in fundamental frequency or passband," *J. Acoust. Soc. Am.* **112**, 2074–2085.
- Shamma, S. A., Elhilali, M., and Micheyl, C. (2011). "Temporal coherence and attention in auditory scene analysis," *Trends Neurosci.* **34**, 114–123.
- Turgeon, M., Bregman, A. S., and Ahad, P. A. (2002). "Rhythmic masking release: Contribution of cues for perceptual organization to the cross-spectral fusion of concurrent narrow-band noises," *J. Acoust. Soc. Am.* **111**, 1819–1831.
- Turgeon, M., Bregman, A. S., and Roberts, B. (2005). "Rhythmic masking release: Effects of asynchrony, temporal overlap, harmonic relations, and source separation on cross-spectral grouping," *J. Exp. Psychol. Hum. Percept. Perform.* **31**, 939–953.
- van de Par, S., and Kohlrausch, A. (1998). "Analytical expressions for the envelope correlation of narrow-band stimuli used in CMR and BMLD research," *J. Acoust. Soc. Am.* **103**, 3605–3620.
- van Noorden, L. P. A. S. (1975). "Temporal coherence in the perception of tone sequences," Ph.D. dissertation, Institute for Perception Research, Eindhoven, The Netherlands.
- Verhey, J. L., Dau, T., and Kollmeier, B. (1999). "Within-channel cues in comodulation masking release (CMR): Experiments and model prediction using a modulation filter bank model," *J. Acoust. Soc. Am.* **106**, 2733–2745.
- Verhey, J. L., Ernst, S. M. A., and Yasin, I. (2012). "Effects of sequential streaming on auditory masking using psychoacoustics and auditory evoked potentials," *Hear. Res.* **285**, 77–85.
- Vliegen, J., Moore, B. C. J., and Oxenham, A. J. (1999). "The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task," *J. Acoust. Soc. Am.* **106**, 938–945.
- Vliegen, J., and Oxenham, A. J. (1999). "Sequential stream segregation in the absence of spectral cues," *J. Acoust. Soc. Am.* **105**, 339–346.
- Zera, J., and Green, D. M. (1993a). "Detecting temporal asynchrony with asynchronous standards," *J. Acoust. Soc. Am.* **93**, 1571–1579.
- Zera, J., and Green, D. M. (1993b). "Detecting temporal onset and offset asynchrony in multicomponent complexes," *J. Acoust. Soc. Am.* **93**, 1038–1052.
- Zera, J., and Green, D. M. (1995). "Effect of signal component phase on asynchrony discrimination," *J. Acoust. Soc. Am.* **98**, 817–827.