Auditory Stream Segregation and the Perception of Across-Frequency Synchrony

Christophe Micheyl, Cynthia Hunter, and Andrew J. Oxenham
University of Minnesota

This study explored the extent to which sequential auditory grouping affects the perception of temporal synchrony. In Experiment 1, listeners discriminated between 2 pairs of asynchronous "target" tones at different frequencies, A and B, in which the B tone either led or lagged. Thresholds were markedly higher when the target tones were temporally surrounded by "captor tones" at the A frequency than when the captor tones were absent or at a remote frequency. Experiment 2 extended these findings to asynchrony detection, revealing that the perception of synchrony, one of the most potent cues for simultaneous auditory grouping, is not immune to competing effects of sequential grouping. Experiment 3 examined the influence of ear separation on the interactions between sequential and simultaneous grouping cues. The results showed that, although ear separation could facilitate perceptual segregation and impair asynchrony detection, it did not prevent the perceptual integration of simultaneous sounds.

Keywords: perceptual organization, auditory perception, stream segregation, asynchrony

Sensitivity to seemingly low-level sensory features in a scene can be profoundly influenced by the way in which the scene is perceptually organized. A striking example of this in visual perception relates to the finding by Ringach and Shapley (1996) that angular discrimination accuracy for simple geometric figures ("pac-men") depends dramatically on whether the figures evoke illusory contours. Effects such as this provide a way for psychophysicists to measure perceptual organization "objectively," on the basis of participants' performance, rather than relying on subjective descriptions or ratings.

In the auditory modality, one of the most compelling examples of the influence of perceptual organization on sensitivity comes from demonstrations that listeners are largely unable to correctly perceive the relative timing of sounds that form part of separate perceptual "streams." The first lines of evidence for this were provided by Broadbent and Ladefoged (1959), Warren, Obusek, Farmer, and Warren (1969), and Bregman and Campbell (1971). These authors showed that listeners cannot correctly tell in which temporal order sounds having different timbres (e.g., a tone, a noise, and a vowel), or tones at different frequencies, are played. Bregman and Campbell (1971) interpreted these results in terms of "auditory streaming," or the perceptual organization of sound

sequences, and concluded that it was difficult to judge the temporal relationship between two sound events when they fell into different perceptual streams. The phenomenon of auditory streaming has been extensively studied with sequences of pure tones that alternate between two frequencies, A and B, forming a repeating ABAB or ABA ABA pattern (Miller & Heise, 1950; van Noorden, 1975, 1977). When the frequency difference between the A and B tones is relatively small, listeners usually perceive the repeating sequence as a single, coherent stream. However, as the frequency separation between the A and B tones increases and the intertone interval decreases, the A and B tones are more likely to split perceptually into separate streams. Auditory stream segregation has also been observed with various other types of stimuli, including complex tones or noises with different spectral or temporal characteristics (Bregman, Ahad, & Van Loon, 2001; Cusack & Roberts, 1999, 2000, 2004; Grimault, Bacon, & Micheyl, 2002; Grimault, Micheyl, Carlyon, Arthaud, & Collet, 2000; Hartmann & Johnson, 1991; Roberts, Glasberg, & Moore, 2002, 2008; Vliegen & Oxenham, 1999; see also reviews in Bregman, 1990; Moore & Gockel, 2002), and is thought to play an important role in speech perception in complex backgrounds (e.g., Oxenham, 2008). Furthermore, auditory streaming has been found in multiple animal species (for a review, see Bee & Micheyl, 2008), suggesting that the phenomenon is quite general.

Following Broadbent and Ladefoged's (1959) and Bregman and Campbell's (1971) seminal findings, several studies have provided further evidence that auditory streaming influences sensitivity to temporal relationships. For instance, Vliegen, Moore, and Oxenham (1999) measured thresholds for the discrimination of changes in timing between two complex tones (A and B) played in an alternating (ABA-ABA) pattern. They found that sensitivity to asynchronies between the A and B tones was impaired in conditions where there was a large fundamental frequency difference between the A and B tones, which led to the formation of separate perceptual streams (Vliegen & Oxenham, 1999). Similar conclu-

sions were drawn by Roberts et al. (2002), who showed that temporal discrimination was impaired across auditory streams, even when the streams were formed of harmonic complex tones having identical power spectra but different temporal envelopes, yielding a difference in perceived pitch.

In all the studies discussed so far, none of the sounds overlapped in time. Although this may limit the generality of the findings somewhat, there is no strong reason to suspect that the conclusions of these studies would not generalize to sounds that overlap partially in time. However, there are several reasons to expect that, in the particular case where the A and B tones begin and end at exactly the same time, the results may be different.

The first reason is that there is some evidence that synchrony detection involves mechanisms that are different from those involved in the discrimination of temporal order (Mossbridge, Fitzgerald, O'Connor, & Wright, 2006) or temporal interval durations (Wright, Buonomano, Mahncke, & Merzenich, 1997). Synchrony detection can be achieved with neural coincidence detectors, which, as their name indicates, are neurons specifically sensitive to concurrent synaptic inputs and are found at very early stages of the auditory pathways (e.g., Oertel, Bal, Gardner, Smith, & Joris, 2000). Thus, discriminating a synchronous from an asynchronous pair of tones could be achieved simply by deciding which stimulus produced the higher response from across-frequency coincidence detectors. In contrast, discriminating the duration of temporal intervals between consecutive sounds probably requires mechanisms that measure and compare elapsed times between neural events over tens of milliseconds or more. These different neural mechanisms may take place at different stages of processing in the auditory system and may be affected differently by sequential grouping mechanisms; see Oxenham (2000) for a similar argument regarding the difference between temporal gap detection and discrimination.

A second reason why synchrony detection might not be affected as much by sequential grouping as temporal interval or order discrimination is that listeners are in general much more sensitive to synchrony than they are to changes in interval duration or temporal order. For instance, Zera and Green (1993a, 1993b, 1993c) measured thresholds for the detection of an asynchrony between frequency components in a complex tone; in some conditions, thresholds were less than 1 ms. In contrast, thresholds for temporal-order identification typically exceed 10 ms (e.g., Hirsch, 1959; Pastore, Harris, & Kaplan, 1982), unless the stimuli are very brief (<5–10 ms), in which case performance may no longer be based on explicit identification of the temporal order (Warren, 2008). In auditory streaming experiments, the tones are usually longer than 25 ms. The tones used by Roberts et al. (2002) were 60 ms, and, even in conditions where listeners clearly heard a single stream, temporal discrimination thresholds were in excess of 10 ms. Vliegen and Oxenham (1999) measured somewhat lower temporal discrimination thresholds in musically experienced listeners, but, across all listeners, the average threshold in the "easiest" condition was still around 10 ms. These data suggest that synchrony detection thresholds can be approximately 5–10 times lower than temporal order or interval discrimination thresholds. In view of this difference, it is possible that listeners might still be able to detect synchrony reliably in conditions where sequential grouping leads to a reduction in the ability to discriminate temporal order.

Finally, there are results in the literature that demonstrate interactions between sequential grouping and synchrony-based simultaneous grouping (e.g., Bregman, 1978; Bregman & Pinker, 1978; Carlyon, 1994; Darwin & Hukin, 1997; Darwin, Hukin, & al-Khatib, 1995; Darwin, Pattison, & Gardner, 1989; Hukin & Darwin, 1995; Shinn-Cunningham, Lee, & Oxenham, 2007). For instance, Darwin et al. (1995) showed that a sequence of precursor tones at the same frequency as a "target" component within a complex tone can lead to perceptual "capture" of the target into the sequence, even when the target is gated simultaneously with (and is harmonically related to) the other frequency components of the complex tone; the contribution of the target tone to the timbre and pitch of the complex tone is then reduced or lost. On the other hand, synchronous onsets (along with congruent spatial and harmonic relations) between a target and captor tones can alter the perception of sequential streams (Shinn-Cunningham et al., 2007), suggesting that synchrony may continue to play a role in perceptual organization, even in the presence of competing sequential cues.

The main aim of the present study was to assess whether, and to what extent, sequential grouping influences the perception of timing between synchronous or near-synchronous sounds.

Experiment 1

The first experiment measured listeners' sensitivity to the direction of temporal asynchrony between a pair of target tones, either in isolation or preceded by a tonal sequence that was designed to perceptually "capture" one of the target into a sequential stream. The design of the experiment was similar to that of previous studies that have examined the influence of sequential grouping on listeners' ability to discriminate temporal order of nonoverlapping tones (Bregman & Campbell, 1971; O'Connor & Sutter, 2000; Sutter, Petkov, Baynes, & O'Connor, 2000). The question was whether sequential streaming would impair listeners' ability to make accurate judgments of the relative temporal onsets of the two near-synchronous and overlapping target tones.

Method

Participants.   Three females and 3 males between the ages of 17 and 41 years served as participants (all but 1 participant were under the age of 25). Before the experiment, participants were tested for normal hearing with a Madsen Conera Audiometer and TDH-49 headphones. All had normal hearing, defined as pure-tone hearing thresholds of 15 dB HL or less at audiometric (octave) frequencies between 250 and 8000 Hz. Two of the listeners had musical training. All participants provided written informed consent and were compensated at an hourly rate.

Stimuli.   All stimuli were pure tones with total durations of 100 ms, including 10-ms raised-cosine onset and offset ramps to avoid audible "spectral splatter." The tones were presented at a level of 60 dB SPL each. A schematic diagram of the various stimulus conditions tested in this experiment is shown in Figure 1. The first condition (upper panel) involved only two target tones: Tone A, the frequency of which was constant at 1000 Hz, and Tone B, the frequency of which was Δf semitones higher. Two Δf conditions were tested: In one, Δf was equal to 6 semitones, resulting in the B tone frequency being set to 1414.2 Hz; in the
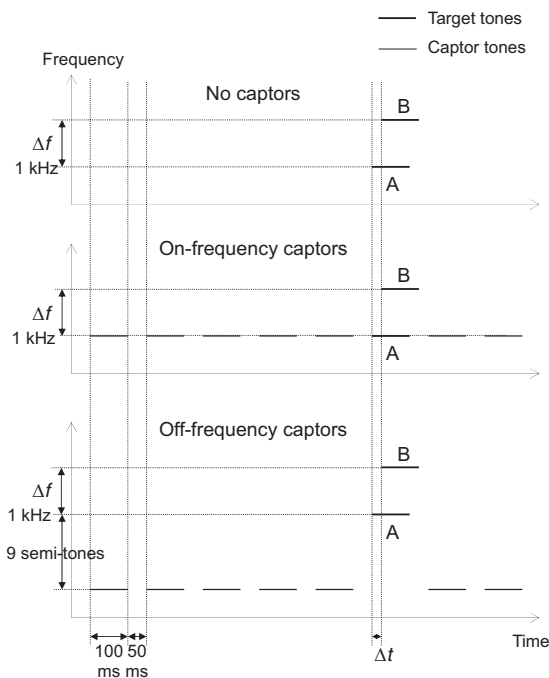
Figure 1. Schematic spectrogram of the stimuli in Experiment 1. The target A and B tones are indicated by letters. The delay between the A and B tones, Δt, was the tracking variable. Only the case in which the B tone lagged the A tone is illustrated in the figure. The A-tone frequency was fixed at 1 kHz. Two frequency separations between the A and B tones, Δf, were tested: 6 and 9 semitones. Depending on the condition, the target tones were presented in isolation (no-captor condition, shown in the upper panel) or preceded and followed by captor tones, the frequency of which was either the same frequency as the target A tone (on-frequency captor condition, shown in the middle panel) or nine semitones lower (off-frequency captor condition, shown in the lower panel).

other, it was equal to 15 semitones, yielding a frequency of 2378.4 Hz for the B tone. These tones usually overlapped in time but were never exactly synchronous. Figure 1 illustrates the case in which the onset of the A tone preceded that of the B tone.

In the other two conditions, the target A tone was preceded and followed by five and two "captor" tones, respectively. These tones were separated from each other, and from the target A tone, by 50-ms silent gaps. Depending on the condition being tested, the frequency of the captor tones was either the same as that of the target A tone (1000 Hz; middle panel in Figure 1), or 9 semitones lower (i.e., 594.6 Hz; bottom panel in Figure 1). For brevity, we hereinafter refer to the former condition as the *on-frequency* condition and to the latter condition as the *off-frequency* condition; the condition involving no captor tones is referred to as the *no-captor* condition.

Procedure. Each trial consisted of two intervals, separated by a 500-ms silent gap. In one of the two intervals, chosen at random to be the first or the second one with equal probability, the B tone lagged the A tone by a time delay, Δt, as illustrated in Figure 1; in the other interval, the B tone preceded the A tone by the same delay. Listeners were instructed to select the interval in which the A tone preceded the B tone. After each trial, feedback was provided in the form of a message ("correct" or "wrong") displayed on the computer screen.

We measured thresholds using an adaptive staircase procedure with a 3-down 1-up rule, which tracks the point corresponding to 79.4% correct on the psychometric function (Levitt, 1971). At the beginning of each "run" through the adaptive procedure, Δt was set to 50 ms. Initially, after three consecutive correct responses or one incorrect response, the value of Δt was decreased or increased, respectively, by a factor of 4. This step size was reduced to a factor of 2 after the first negative-going reversal (or "peak") in the adaptive staircase and to 1.414 (approximately $\sqrt{2}$) after two more reversals. A run was terminated after a further six reversals in the adaptive procedure, and the threshold for the run was computed as the geometric mean of Δt at the last six reversal points. During each run, the value of Δt was not allowed to exceed 100 ms. If the adaptive staircase procedure called for a value of Δt greater than 100 ms, Δt was set to 100 ms, and the procedure continued.

At least four runs per condition were completed by each participant. Four of the 6 participants actually performed between 8 and 12 runs in each condition because we wanted to test for possible learning effects. As thresholds were fairly stable across repetitions, geometric mean thresholds were calculated across all the runs collected in each condition in a given participant. The conditions were run in an order that was selected at random for each participant and each repetition. Participants were tested individually in 2-hr sessions. On average, the experiment took approximately three sessions (6 hr) per participant to complete.

Apparatus. The stimuli were generated in Matlab and were presented by means of a 24-bit digital-to-analog converter (LynxStudio Lynx22) to both earphones of a Sennheiser HD580 headset. Listeners were seated in a sound-attenuating double-walled chamber, and they responded to stimuli through a computer keyboard or mouse. Intervals were marked by lights on a virtual response box presented on a computer monitor in the booth.

Data analysis. Consistent with the use of a geometric tracking rule in the adaptive procedure, geometric (rather than arithmetic) means were used when averaging thresholds across repetitions or listeners, and statistical analyses were performed on log-transformed thresholds (rather than threshold in linear, ms units). Statistical analyses consisted mainly of repeated measures analyses of variance (RM-ANOVAs). Reported *p* values and degrees of freedom include the Greenhouse–Geisser correction wherever required (i.e., whenever the sphericity assumption was not met). All statistical analyses were performed with SPSS Version 16 (SPSS Inc.).

Results

The results of this experiment, averaged across participants, are shown in Figure 2, which plots the threshold value of Δt as a function of the frequency separation between the two tones, Δf. A significant main effect of captor condition was observed, $F(1.06, 5.32) = 9.56$, $p = .024$, $\eta^2 = 0.657$; thresholds were significantly higher in the on-frequency captor condition than in the no-captor condition, $F(1, 5) = 12.19$, $p = .017$, $\eta^2 = 0.709$; and in the off-frequency captor condition, $F(1, 5) = 7.25$, $p = .043$, $\eta^2 = 0.592$. Note that the mean thresholds measured in the on-frequency captor condition are relatively close to 100 ms, the largest Δt value allowed in the tracking procedure. In fact, all listeners occasionally obtained thresholds larger than 50 ms in these on-frequency captor conditions. Considering that 50 ms corresponds to just two $\sqrt{2}$
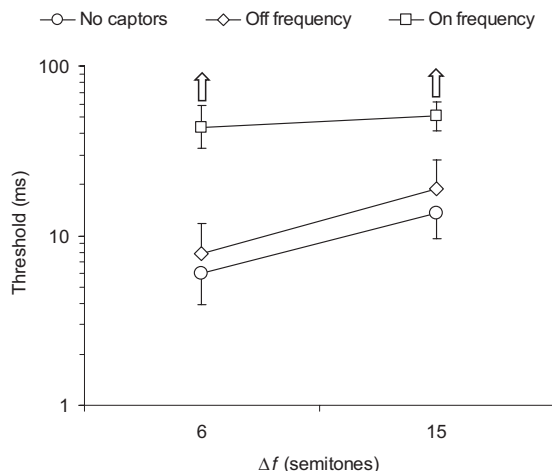
Figure 2. Results of Experiment 1: Influence of on- and off-frequency captor tones on thresholds for the discrimination of the direction of an asynchrony. Each data point represents a mean threshold across 6 listeners. The error bars reflect interparticipant standard errors; upper or lower bars are omitted wherever necessary to avoid overlap. The different symbols denote different captor conditions, as indicated in the legend at the top of the plot: Circles represent no captors, diamonds represent off-frequency captors, and squares represent on-frequency captors. The frequency separation, Δf, between the two target tones is indicated on the *x*-axis; it is expressed in semitones (1 semitone is 1/12th of an octave, or roughly 6%).

steps down from 100 ms on the geometric tracking scale, we cannot rule out the possibility that the mean thresholds obtained in the 6- and 15-semitone on-frequency captor conditions (approximately 43 and 50 ms, respectively) represent ceiling; therefore, the statistical difference between the on-frequency condition and the others may, if anything, be an underestimation of the influence of the on-frequency captors on listeners' ability to perceive the relative timing of the two target tones. Upward-pointing arrows are shown for the appropriate conditions in Figure 2 to emphasize this point.

In addition to the large effect of the on-frequency captors, a much smaller but nonetheless significant difference was observed between the no-captor and off-frequency captor conditions, $F(1, 5) = 8.66$, $p = .032$, $\eta^2 = 0.634$; with slightly higher (poorer) thresholds in the off-frequency captor condition than in the no-captor condition.

A significant main effect of frequency separation was also found, $F(1, 5) = 24.84$, $p = .004$, $\eta^2 = 0.832$. As indicated by a significant interaction of captor condition and Δf, $F(1.09, 5.44) = 10.12$, $p = .021$, $\eta^2 = 0.669$; the influence of Δf depended on the condition being tested: It was significant in the no-captor condition, $F(1, 5) = 14.19$, $p = .013$, $\eta^2 = 0.739$; and in the off-frequency captor condition, $F(1, 5) = 41.73$, $p = .001$, $\eta^2 = 0.893$; but not in the on-frequency captor condition, $F(1, 5) = 1.55$, $p = .268$, $\eta^2 = 0.237$; which may again reflect the fact that thresholds in the on-frequency captor condition approached ceiling at both values of Δf.

Discussion

The results are consistent with the hypothesis that stream segregation adversely affects listeners' ability to make accurate judg-

ments of relative timing between simultaneously present sounds. The on-frequency captor tones led to highly elevated discrimination thresholds, presumably because the captors resulted in the A and B tones being perceived within separate auditory streams, which in turn led to an impaired ability to make temporal judgments between them. The influence of the off-frequency captor tones was much less than that of the on-frequency captors, suggesting that the main effect cannot be accounted for by a "distraction" effect caused by the presence of any interfering sound.

A slight complication in the interpretation of the thresholds measured in the on-frequency-captors condition relates to the fact that, when Δt exceeded 50 ms (in either direction), the B tone started to overlap temporally with the preceding (or following) captor tone. It is not clear whether and how performance was affected by this partial overlap. In any case, this does not affect the conclusion that thresholds were systematically higher in the on-frequency condition than in the other two conditions.

Overall, these findings complement those of earlier studies that used strictly sequential (i.e., temporally nonoverlapping) sounds (Bregman & Campbell, 1971; Hong & Turner, 2006; O'Connor & Sutter, 2000; Roberts et al., 2002, 2008; Vliegen et al., 1999). They show that the main conclusion of these studies—that the ability to perceive fine temporal relationships between individual sounds in a sequence depends on how these sounds are perceptually organized—also applies to sounds that are nearly synchronous and overlap in time.

Experiment 2

The results from Experiment 1 confirm earlier findings of a reduced ability to make accurate timing judgments between sound elements in different streams (Bregman & Campbell, 1971; Broadbent & Ladefoged, 1959; Roberts et al., 2002; Vliegen et al., 1999), and they reveal that this effect is present, even when the two elements temporally overlap and are nearly synchronous. However, the target tones in that Experiment 1 were never quite synchronous. As mentioned in the introduction, there are several reasons to suspect that the situation might be different for synchronous events. In particular, there is some evidence that synchrony detection involves mechanisms that are different from those involved in the discrimination of temporal order (Mossbridge et al., 2006). Discriminating an A-leading pair from a B-leading pair (as in Experiment 1) appears to require a more explicit temporal judgment, which may be more susceptible to the constraints of auditory sequential streaming. Discriminating a synchronous from an asynchronous pair of tones can be achieved simply by deciding which stimulus produced the higher response from across-frequency coincidence detectors. An analogy can be made here with experiments that measure participants' ability to detect a change in an ongoing stimulus: In some cases, people are more accurate at detecting a change than they are at discriminating the direction of the change (e.g., Macmillan, 1971, 1973; for a review, see Micheyl, Kaernbach, & Demany, 2008), and the pattern of results suggests that discrimination and detection may be subserved by distinct mechanisms (see also Hafter, Bonnel, & Gallun, 1998). Indeed, thresholds for discriminating the direction of a temporal shift between two tones (Hirsch, 1959; Pastore et al., 1982) are often considerably larger than thresholds for the detection of synchrony (Zera & Green, 1993a, 1993b, 1993c). It may be

that, unlike the explicit temporal comparisons required in Experiment 1, synchrony detection is governed by low-level auditory mechanisms, which are not affected by perceptual organization. Alternatively, synchrony detection might be affected by perceptual organization, either in the same way or to a lesser extent than temporal order discrimination.

To address this question, in Experiment 2 we measured thresholds for the detection of an asynchrony between target tones in a two-interval, two-alternative task in which one of the two intervals always contained a completely synchronous tone pair and the other interval always contained an asynchronous pair.

Method

Participants.   Eight listeners took part in this experiment. Four of them also participated in Experiment 1. Because of limited availability, 2 of the 6 listeners who took part in Experiment 1 could not be included into Experiment 2. To increase the sample size for this experiment, we added 4 new listeners (2 male, 2 female), with ages ranging from 19 to 28. These participants also had normal hearing, as defined in Experiment 1. One of the new listeners was the second author (Cynthia Hunter), who did not receive payment for her participation as a listener.

Stimuli and procedure.   The basic stimuli (A and B target tones and captors) were the same as in the previous experiment. The main difference between the two experiments is that, whereas in the previous experiment the B tone was always shifted in time (forward or backward) relative to the target A tone, in this experiment, the B tone was always synchronous with the target A tone in one of the two observation intervals in a trial and was asynchronous in the other interval. This time, two adaptive tracks were randomly interleaved: In one track, the B tone was shifted backward relative to the A tone in the asynchronous interval; in the other track, the B tone was shifted forward. As in the previous experiment, the temporal shift, Δt, applied to the B tone was varied adaptively using a three-down, one-up rule. The specifics of the procedure, including the step sizes and termination rules, were identical to those used in Experiment 1. Note that the difference in the timing of the B tones (relative to the A tone) between the two intervals in Experiment 1 was $2\Delta t$, whereas in this experiment it was just Δt. For this reason, the starting value of Δt was set to the maximum value of 100 ms, instead of the 50 ms used in Experiment 1.

Results

The results of Experiment 2 are shown in Figure 3. There was a significant main effect of captor condition, $F(1.12, 7.82) = 24.39$, $p = .001$, $\eta^2 = 0.777$. Thresholds were significantly larger in the on-frequency captor condition than in the no-captor condition, $F(1, 7) = 33.36$, $p = .001$, $\eta^2 = 0.827$; or the off-frequency captor condition, $F(1, 7) = 20.62$, $p = .003$, $\eta^2 = 0.747$. As in the previous experiment, listeners often obtained thresholds in excess of 70 ms on one or more runs in the on-frequency captor condition at the 15-semitone separation, making it possible that the mean threshold measured in that condition (around 44 ms) would have been higher if a larger maximum Δt had been allowed in the experiment. Thresholds did not differ significantly between the off-frequency captor condition and the no-captor condition, $F(1, 7) = 0.52$, $p = .492$, $\eta^2 = 0.070$.
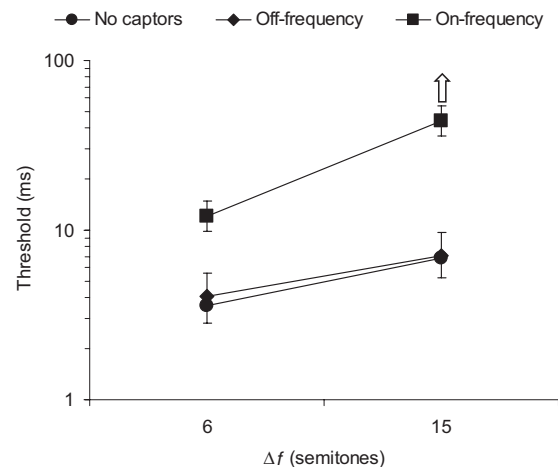

Figure 3.   Results of Experiment 2: Influence of on- and off-frequency captor tones on thresholds for the detection of an asynchrony. Each data point represents a mean threshold across 8 listeners. The error bars reflect interparticipant standard errors; upper or lower bars are omitted wherever necessary to avoid overlap. The different symbols denote different captor conditions, as indicated in the legend at the top of the plot: Circles represent no captors, diamonds represent off-frequency captors, squares represent on-frequency captors. The frequency separation in semitones, Δf, between the two target tones is shown on the x-axis.

In addition to the effect of captor condition, there was a significant main effect of Δf, $F(1, 7) = 77.44$, $p < .0005$, $\eta^2 = 0.917$; and a significant interaction of captor condition and Δf, $F(1.62, 11.37) = 4.80$, $p = .026$, $\eta^2 = 0.407$. Further analysis revealed that the effect of Δf was significantly larger in the on-frequency captor condition than in the other two captor conditions, $F(1, 7) = 6.44$, $p = .039$, $\eta^2 = 0.479$.

Finally, no significant difference was observed between thresholds estimated on the basis of trials in which the B tone was shifted forward relative to the A tone in the asynchronous interval, and thresholds estimated on the basis of trials in which the shift was in the opposite direction, $F(1, 7) = 0.028$, $p = .873$, $\eta^2 = 0.004$. This finding justifies our averaging of these thresholds in Figure 3.

Discussion

The results of this experiment are consistent with the results from Experiment 1 and support the hypothesis that sequential stream segregation, on the basis of spectral and temporal proximity, renders temporal judgments between streams inaccurate. The novel finding here is that the deterioration in temporal judgments between streams is found even when the two target sounds are synchronous. This reveals that synchrony detection is not immune from sequential grouping influences: Listeners' sensitivity to synchrony depends upon sequential grouping, just like sensitivity to other forms of temporal shifts between tones (e.g., Bregman & Campbell, 1971; Broadbent & Ladefoged, 1959; Roberts et al., 2002, 2008; Vliegen et al., 1999). Thus, although there is evidence for the existence of "coincidence detectors" or "synchrony detectors" in the auditory system (Oertel et al., 2000), the present findings suggest that listeners' conscious access to the output of these detectors is limited by perceptual organization processes.

A comparison of the results from Experiments 1 and 2 in the no-captor conditions can provide some insight into the question of whether synchrony detection and temporal order discrimination are governed by the same mechanisms. In principle, both could be achieved by a direct comparison of the onset (or offset) times of the target A and B tones in each interval. Under such circumstances, a signal-detection-theoretic model could be applied on the basis of the traditional assumptions of statistically independent sensory observations contaminated by equal-variance Gaussian noise. Such a model predicts that an ideal observer could achieve a threshold Δt approximately 2.91 times *smaller* in the asynchrony direction identification task (Experiment 1) than in the asynchrony detection task (Experiment 2).[1] In fact, as shown in Figure 4, the 4 listeners who took part in both experiments obtained significantly *larger* thresholds (by a factor of 1.84) in the no-captor and off-frequency conditions of Experiment 1 than in Experiment 2, $F(1, 3) = 17.83$, $p = .024$, $\eta^2 = 0.856$. This result supports the conclusion of Mossbridge et al. (2006), arrived at with different means, that asynchrony detection and asynchrony order discrimination may be based on at least partly different cues. Consistent with this explanation, thresholds increased significantly with Δf in the on-frequency captor condition in Experiment 2, $F(1, 3) = 56.44$, $p < .0005$, $\eta^2 = 0.89$; but not in Experiment 1. However, given the potential for ceiling effects in Experiment 1, we cannot rule out the possibility that a significant effect of Δf would have been observed in the on-frequency captor condition of that experiment if thresholds had been lower overall.

The increase in threshold with increasing Δf in the on-frequency captor condition may be related to the greater overlap in the peripheral representations of the two tones at the smaller frequency separation. The tonotopic organization (or frequency-to-place mapping) that occurs in the cochlea produces some overlap between sounds that are close in frequency. It is possible that some locations along the cochlea's basilar membrane responded to both A and B tones at the 6-semitone separation. In Experiment 2, such locations would have registered a single onset response in the synchronous tone intervals but a double onset response in the asynchronous intervals. This difference may have been easier to distinguish (leading to lower thresholds) than the two double onset responses that would have occurred in the case of the asynchrony discrimination in Experiment 1. This explanation is explored further in Experiment 3B.

Experiment 3

The results of Experiments 1 and 2 indicate that sequential stream segregation results in a reduced ability to detect and discriminate onset and offset synchrony cues. Previous studies have found that sounds presented to different ears tend to be segregated perceptually, and sounds presented to the same ear tend be grouped together, at least in the absence of other competing cues (for a review, see Bregman, 1990). However, in situations in which other grouping cues are present, "ear of entry" cues can be trumped by other cues. A well-known example of this is Deutsch's "octave illusion" (Deutsch, 1974, 2004), in which tones at two different frequencies (A and B) that are presented dichotically in alternating fashion (ABAB in one ear vs. BABA in the other ear) are erroneously perceived as having a constant frequency in each ear. In that example, frequency similarity of the tones over time overrides the very strong spatial location cues produced by presenting simultaneous sounds to opposite ears.

Experiments 3A–3C were designed to assess the relative influence of sequential grouping and the ear of presentation on the detection of across-frequency synchrony. In particular, we were interested in determining whether presenting the target A tone in the opposite ear to the on-frequency captor tones was enough to eliminate the effect of the on-frequency captors on thresholds or whether sequential grouping based on frequency proximity was
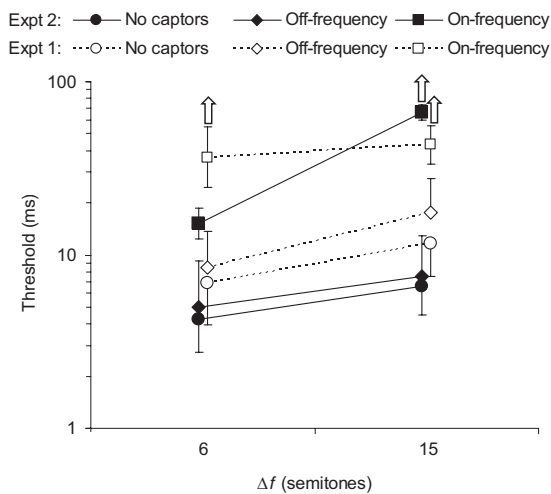

Figure 4. Comparison between Experiment 1 and Experiment 2 results: Difference between thresholds for asynchrony direction discrimination (Experiment 1) and asynchrony detection (Experiment 2). Each data point represents the mean threshold across the 4 listeners who took part in both experiments. The error bars reflect interparticipant standard errors. The different symbols denote different captor conditions, as indicated in the legend at the top of the plot: Circles represent no captors, diamonds represent off-frequency captors, and squares represent on-frequency captors. Open symbols correspond to Experiment 1; filled symbols correspond to Experiment 2. The frequency separation in semitones, Δf, between the two target tones is indicated on the *x*-axis.

---

[1] The prediction of thresholds 2.91 times smaller in Experiment 1 (asynchrony direction identification) than in Experiment 2 (asynchrony detection) derives from a signal-detection-theoretic model based on the assumption that the onset times of the A and B tones in each observation interval are represented internally as independent Gaussian random variables with equal variance and an expected value related linearly to onset time. Following an analysis similar to that described in Micheyl et al. (2008), it can be shown that, for an optimal (likelihood ratio) observer, the relationship between d' and the proportion of correct responses, Pc, is given by

$$d' = \Phi^{-1}(Pc) \tag{1}$$

in Experiment 1 and

$$d' = 2\Phi^{-1}\left(\frac{1}{2} + \sqrt{\frac{Pc}{2} - \frac{1}{4}}\right) \tag{2}$$

in Experiment 2. In these equations, Φ^{-1} denotes the inverse cumulative standard normal distribution. Substituting 0.794 for Pc in Equations 1 and 2, and taking the ratio, one obtains the value of 2.91 cited in the main text.

sufficiently powerful to hinder synchrony detection, despite the incongruent spatial information between the target and the captors.

Method

Participants. Each of the three parts of this experiment was completed by 4 listeners. In Experiments 3A and 3B, the 4 listeners had also taken part in Experiment 2. Three of these listeners also took part in Experiment 3C. The fourth listener could not complete Experiment 3C because of schedule conflicts. This listener was replaced by another one, who had already taken part in Experiment 1 and was available to take part in Experiment 3C.

Stimuli and procedure. The stimuli used in Experiments 3A–3C are illustrated in Figure 5. In Experiment 3A (upper panel), the captor tones were presented to the left ear and the two target tones (A and B) were presented to the right ear. In Experiment 3B (middle panel), the target A tone and the captor tones were presented to the left ear; only the B tone was presented to the right ear. In Experiment 3C (lower panel), the B tone was presented to the left ear, together with the captor tones, and the A tone was presented to the right ear.

All other stimulus parameters were identical to those used in Experiment 2, as was the procedure, which required participants to distinguish a completely synchronous tone pair from a pair that was asynchronous, with a time difference of Δt.

Results and Discussion

For brevity, only the most relevant effects are detailed here. In none of these three experiments were thresholds found to differ
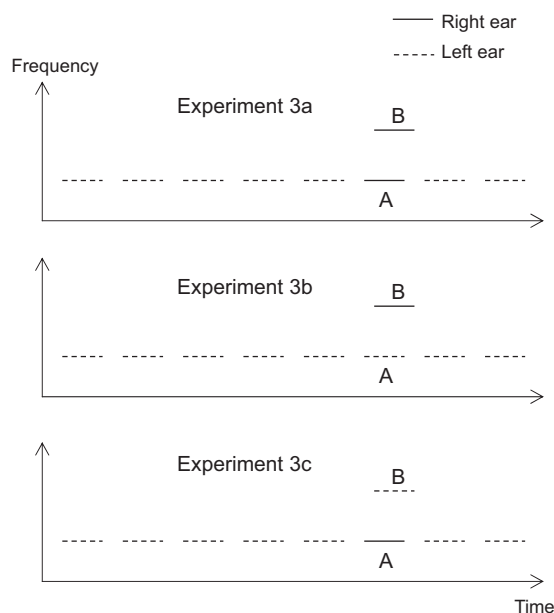

Figure 5. Schematic spectrograms of the stimuli in Experiments 3A–3C. In Experiment 3A (upper panel), the target A and B tones were presented to the right ear and the captor tones were presented to the left ear. In Experiment 3B (middle panel), the target B tone was presented to the right ear and the target A tone and the captor tones were presented to the left ear. In Experiment 3C (lower panel), the target A tone was presented to the right ear and the target B tone and the captor tones were presented into the left ear.

significantly depending on whether the B tone was advanced or delayed relative to the A tone in the asynchronous interval; accordingly, the thresholds from the two adaptive tracks were once again averaged together.

Experiment 3A. The results of Experiment 3A are shown in Figure 6A. The filled symbols correspond to the mean thresholds measured in this experiment. To facilitate comparisons with Experiment 2, the mean thresholds measured in the same 4 listeners, which were plotted as filled symbols in Figure 3, are replotted as open symbols.

For the no-captor condition and the off-frequency captor condition, the thresholds measured in this experiment were not statistically different from those measured in Experiment 2, $F(1, 3) = 0.06$, $p = .945$, $\eta^2 = 0.002$. In contrast, thresholds in the on-frequency captor condition were markedly lower here than in the Experiment 2, $F(1, 3) = 26.58$, $p = .014$, $\eta^2 = 0.899$; in fact, they were statistically indistinguishable from those measured in the no-captor condition, $F(1, 3) = 0.61$, $p = .492$, $\eta^2 = 0.169$. The only significant difference across the different conditions of this experiment was a slight but statistically significant threshold elevation in the off-frequency captor condition, compared with the other two conditions, $F(1, 3) = 17.55$, $p = .025$, $\eta^2 = 0.854$.

These results indicate that presenting the captor and target tones to different ears is sufficient to eliminate the effect of sequential grouping on synchrony detection. It is unclear whether this outcome is due to listeners being able to attend selectively to one ear, to perceptual separation of the target and captor tones being promoted by perceived laterality differences, to sequential grouping being "reset" by switching the ear of presentation of the stimuli (Anstis & Saida, 1984; Roberts et al., 2008; Rogers & Bregman, 1998), or to a combination of these factors. In any case, these results show that, although sequential grouping can have a dramatic impact on the ability to detect synchrony, under the test conditions considered here it did not override the effect of ear separation.

The observation that thresholds in the off-frequency captor condition were slightly elevated compared with those measured in the no-captor and on-frequency captor conditions is similar to what was found in Experiment 1 and may be due to some distraction effect, perhaps because attention was drawn exogenously by the off-frequency tones to the opposite ear and different frequency region from the target tones. Nevertheless, this effect is small compared with the much larger effect of the on-frequency captor tones in the previous two experiments.

Experiment 3B. The thresholds measured in this experiment are shown as filled symbols in Figure 6B. As in Figure 6A, the mean thresholds obtained by the same 4 listeners in Experiment 2 are replotted here as open symbols. As in Experiment 2, thresholds in the on-frequency captor condition were significantly elevated compared with those measured in the no-captor and off-frequency captor conditions, $F(1, 3) = 27.30$, $p = .014$, $\eta^2 = 0.901$. Again, the mean thresholds obtained in the on-frequency captor conditions were large enough to suggest a ceiling effect.

It is important to note that the low thresholds in the no-captor and off-frequency captor conditions in this experiment (in which the target tones were in opposite ears), as in Experiment 2 (in which the target tones were in the same ear), indicates that presenting the target tones in different ears did not prevent listeners from judging the relative timing of these tones accurately. Com-
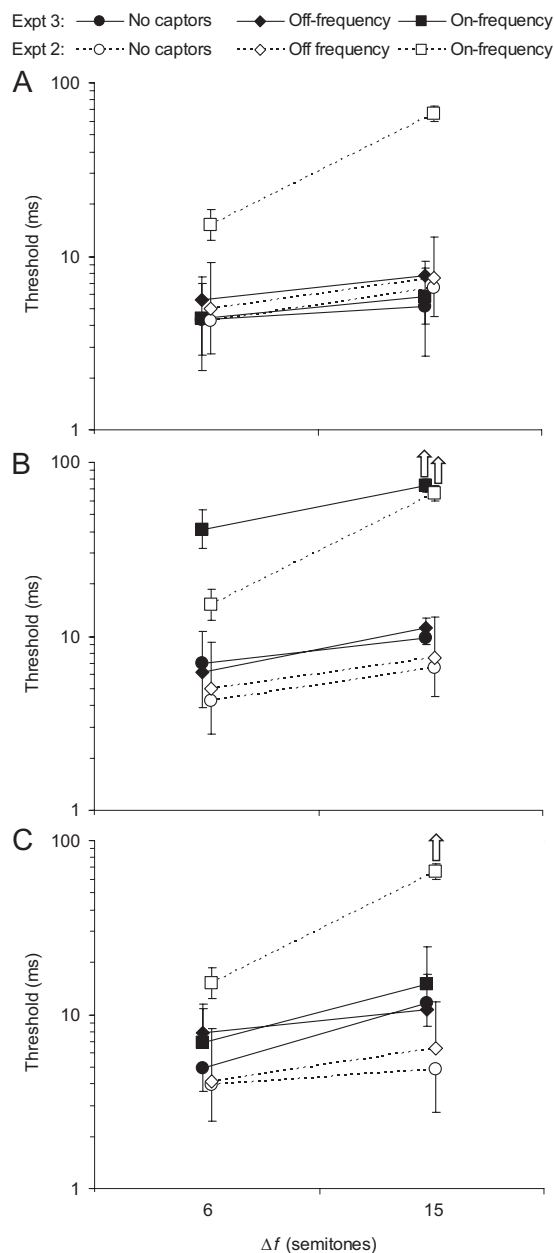
Figure 6. Results of Experiment 3. The different symbols denote different captor conditions, as indicated in the legend at the top of the plot: Circles represent no captors, diamonds represent off-frequency captors, and squares represent on-frequency captors. Filled symbols are used to show thresholds measured in this experiment, averaged across 4 listeners. Open symbols are used to show the mean thresholds obtained by the same 4 listeners in Experiment 2, in which the target and captor tones were presented monaurally into the same ear, and are superimposed to facilitate comparison between the two experiments. The frequency separation, Δf, between the two target tones is indicated underneath the *x*-axis; it is expressed in semitones. The error bars reflect standard errors across listeners. (A) Results of Experiment 3A: Influence on asynchrony detection thresholds of presenting the target and captor tones to opposite ears. Each data point represents a mean threshold across 4 listeners. (B) Results of Experiment 3B: Influence on asynchrony thresholds of presenting the target B tone to the opposite ear as the other tones. (C) Results of Experiment 3C: Influence on asynchrony thresholds of presenting the target A tone to the opposite ear as the other tones.

bined with the finding that the only condition in which listeners were largely unable to detect an asynchrony between the two targets tones was that in which these tones fell into different perceptual streams (the on-frequency captor condition), this suggests that ear separation did not prevent listeners from perceptually integrating the two target tones. This interpretation is consistent with earlier results in the psychoacoustic literature, which also indicate that listeners can integrate different frequency components presented simultaneously into the left and right ears into a single perceived sound (e.g., Bernstein & Oxenham, 2008; Darwin & Ciocca, 1992; Houtsma & Goldstein, 1972). Taken together, the results of the present and previous experiments indicate that, although ear separation can facilitate perceptual segregation, it does not prevent perceptual integration.

Experiment 3C. Of the three ear-of-entry manipulations illustrated in Figure 6, that used in Experiment 3C (shown in Figure 6C) had perhaps the least predictable outcome, because none of the three grouping cues (frequency proximity, synchrony, and ear of entry) were always consistent with each other. No significant difference was observed between thresholds in the on-frequency captor condition and thresholds in the no-captor and off-frequency conditions, $F(1, 3) = 1.70$, $p = .283$, $\eta^2 = 0.283$. The only significant effect was that of Δf, $F(1, 3) = 12.74$, $p = .038$, $\eta^2 = 0.809$. This outcome indicates that separation by ear was sufficient to avoid sequential grouping of the A tone with the on-frequency captor tones and was not sufficient to prevent accurate detection of synchrony between the two target tones.

This experiment was most similar in nature to the illusion of Deutsch (1974), in that one target tone had the same frequency as the captor tones but a different ear of entry, whereas the other target tone shared the same ear of entry as the captor tones but had a different frequency. However, the results are not similar to those found by Deutsch, because in this experiment, frequency proximity did not trump ear of entry in determining the auditory stream. There are many differences between the two experiments that may account for the difference in the results. Perhaps the most important is that Deutsch's experiment involved a sequence formed by several pairs of synchronous tones presented to opposite (and alternating) ears, whereas in the present experiment, only one spatially ambiguous tone pair was presented in a context of several unambiguous single captor tones. It may be that more (or, at least, more extended) ambiguity is required for Deutsch's illusion to be observed.

Another possible reason for the lack of strong capture effect in both this experiment and Experiment 3A, is that switching the ear of presentation between the captor tones and the target A tone led to a resetting of sequential grouping (Anstis & Saida, 1984; Roberts et al., 2008; Rogers & Bregman, 1998). In Experiment 3B and in the on-frequency captors condition of Experiments 1 and 2, such resetting did not occur because the target A tone was always presented in the same ear as the captor tones, so that ear-of-presentation information was consistent throughout the sequence.

Overall, the results of Experiments 2 and 3 are consistent with explanations based on the ear of entry of the stimuli: If the captor tones are at the same frequency and presented to the same ears as one of the target tones, then that target tone is captured into a sequential stream, but if either the frequency or the ear of entry is different, then that target tone is perceived separately from the captor tones and synchrony between it and the other target tone is more accurately perceived. One limitation of the present experimental design is that it

did not allow us to distinguish between ear of entry and perceived location (see Gockel & Carlyon, 1998). Such a distinction could be achieved in future experiments with interaural time differences (ITDs). However, given that ear of entry is generally considered a more potent segregation cue than lateralization based only on ITDs, we predict similar outcomes in such an experiment.

General Discussion

Summary

This study explored the extent to which sequential auditory grouping based on temporal and frequency proximity affects listeners' ability to detect synchrony between simultaneous sounds at different frequencies. The main finding is that sequentially induced segregation significantly hinders listeners' ability to discriminate (Experiment 1) or detect (Experiment 2) temporal asynchrony. The results of a third experiment, which explored the influence of ear of entry (or perceived laterality), showed (a) no disruptive effect of sequential streaming when the target tones were presented together in the opposite ear from that of the captor tones (Experiment 3A); (b) no disruptive effect of sequential streaming when the captors shared the same frequency, but not the same ear, as one of the two target tones; and (c) a significant disruptive effect of sequential streaming when the captors shared the same frequency and the same ear as one of the two target tones. Thus, a difference in the ear of entry (or perceived laterality) is sufficient to hinder the formation of a perceptual stream between a target and the captor tones of the same frequency, even if the spatial location of the target tone is rendered more ambiguous by the synchronous (or near-synchronous) addition of the second target tone in the opposite ear.

Synchrony Detection Is Not Immune From Sequential Grouping Influences

Several earlier results had shown a detrimental influence of sequential organization processes, or auditory streaming, on the perception of the temporal relationships between nonoverlapping sequential sounds (e.g., Roberts et al., 2002; Vliegen et al., 1999). The present results reveal that the impairment extends to the perception of synchrony. This reveals that synchrony detection is not immune from sequential grouping influences: Listeners' sensitivity to synchrony depends upon sequential grouping, just like sensitivity to other forms of temporal shifts between nonoverlapping tones (e.g., Bregman & Campbell, 1971; Broadbent & Ladefoged, 1959; Roberts et al., 2002, 2008; Vliegen et al., 1999).

The results also show that synchrony detection may be used as an indirect measure for perceptual stream segregation in much the same way that timing judgments between nonoverlapping tones have been used in the past (e.g., Roberts et al., 2002, 2008). In a related study, such a measure was used recently to investigate the potential neural correlates of perceptual streaming by using dual streams of synchronous and asynchronous tones at different frequencies to determine whether differences in human perception between repeated synchronous and asynchronous sequences were reflected by differences in responses of single units in the auditory cortex of the awake ferret (Elhilali, Ma, Micheyl, Oxenham, & Shamma, 2009). More generally, objective tasks, such as the ones explored in this study, provide a promising approach to carrying

out simultaneous behavioral and neurophysiological experiments in nonhuman species, for which subjective measures of perception are not usually viable.

As discussed in the introduction, there are several reasons for supposing that synchrony detection involves different mechanisms than temporal-order identification or interval-duration discrimination. Moreover, synchrony detection mechanisms are likely to be much simpler and more "primitive" than those required for temporal order identification or interval-duration discrimination. In principle, synchrony detection could be achieved with an across-frequency coincidence-detection mechanism, which can be implemented at the level of single neurons through rapid integration of synaptic inputs followed by thresholding. Such neural coincidence detectors have been identified in the cochlear nucleus–the first stage of processing after the cochlea (e.g., Oertel et al., 2000). The present findings suggest that listeners do not have conscious access to the outputs of these neural coincidence detectors independent of the formation of auditory streams.

Synchrony Detection and Temporal-Order Identification: Different Mechanisms?

Although the results of this study reveal that stream segregation impairs both synchrony detection and temporal-order identification, this does not imply that these two temporal perception abilities involve similar underlying mechanisms. In fact, the results provide some evidence that the two abilities are at least partly dissociated. This evidence relates to the finding that, in the baseline (no-captors) condition, thresholds in Experiment 2 (asynchrony detection) were considerably smaller than expected based on those measured in Experiment 1 (temporal-order identification): whereas signal detection theory predicts thresholds about 2.91 times *larger* in Experiment 2 than in Experiment 1, the measured thresholds were actually about 1.84 times *smaller* in Experiment 2. Thus, thresholds in a temporal-order identification task were about five times larger than expected based on asynchrony-detection thresholds measured in the same listeners. This large discrepancy between predictions and results suggests that the detection of asynchronies and the identification of temporal order involve different sensory cues and/or mechanisms. A similar conclusion was reached by Mossbridge et al. (2006) based on results in perceptual learning experiments, which showed that learning in a temporal-order identification task did not generalize to performance in an asynchrony detection task, and vice versa, suggesting that the two tasks involve distinct neural mechanisms. As was pointed out earlier, whereas synchrony detection can be achieved simply based on the output of neural coincidence detectors (Oertel et al., 2000), temporal-order identification requires an ability to identify response patterns evoked by two asynchronous tones based on whether the low- or high-frequency tone is lagging, which likely requires more sophisticated neural mechanisms than coincidence detection.

Interactions Between Sequential and Simultaneous Grouping

Several previous studies have shown that sequential grouping can counteract synchrony-based grouping. For instance, Darwin and colleagues have shown in a series of studies (Darwin & Hukin,

1997, 1998; Darwin et al., 1995; Darwin & Sutherland, 1984; Hukin & Darwin, 1995) that presenting preceding tones at the same frequency as one of the components in a synchronous (uniform or vowel-shaped) complex tone substantially reduced the influence of the target harmonic on the perceived pitch, or phonemic identity, of the complex. Further support for a predominance of sequential grouping over synchrony-based grouping was provided more recently by Lee, Shinn-Cunningham, and colleagues (Lee, Babcock, & Shinn-Cunningham, 2008; Lee & Shinn-Cunningham, 2008; Shinn-Cunningham et al., 2007), who also used phonemic identity and perceptual rating tasks to determine the perceptual organization of ambiguous stimuli with competing frequency, synchrony, and spatial cues.

All these studies involved subjective judgments, in which there was no strong incentive for listeners to give a larger weight to synchrony cues than to sequential grouping cues. As such, the results do not address the question of whether attention or simply experimenter instructions can increase the relative weight of synchrony cues over the influence of sequential streaming. There have been a few studies in which the listeners' task was such that it encouraged the dominance of grouping by synchrony over streaming by frequency. Grose and Hall (1993)—and, more recently, Dau, Ewert, and Oxenham (2009)—measured the influence of sequential streaming on comodulation masking release (CMR). The phenomenon of CMR is a release from masking produced by coherent amplitude modulations over time between the masking sound and flanking bands that are synchronous with, but remote in frequency from, the masker. Grose and Hall (1993) and Dau et al. (2009) found that the release from masking is reduced or eliminated when the flanking bands were preceded or followed by a sequence of capture bands designed to perceptually segregate the masker from the flanking bands. These experiments provide an example of where the capture through sequential streaming takes place, despite the use of a task that favors grouping by synchronous onset.

The present experiments provide another objective task, in which performance is enhanced if simultaneous grouping cues can override sequential streaming cues. Importantly, the results show not only that synchrony is not sufficient to bind simultaneous elements when in competition with sequential stream, but also that the perception of synchrony itself is severely impaired. Taken together with the earlier results, these findings indicate that the influence of sequential streaming is "automatic," to the point that it cannot be overcome through changes in attention or task, and that not only perceived organization of an auditory scene, but also sensitivity to basic sound features, is affected by perceptual grouping.

References

Anstis, S., & Saida, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology: Human Perception & Performance, 11,* 257–271.

Bee, M. A., & Micheyl, C. (2008). The cocktail party problem: What is it? How can it be solved? And why should animal behaviorists study it? *Journal of Comparative Psychology, 122,* 235–251.

Bernstein, J. G., & Oxenham, A. J. (2008). Harmonic segregation through mistuning can improve fundamental frequency discrimination. *Journal of the Acoustical Society of America, 124,* 1653–1667.

Bregman, A. S. (1978). Auditory streaming: Competition among alternative organizations. *Perception & Psychophysics, 23,* 391–398.

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

Bregman, A. S., Ahad, P. A., & Van Loon, C. (2001). Stream segregation of narrow-band noise bursts. *Perception & Psychophysics, 63,* 790–797.

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology, 89,* 244–249.

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. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America, 31,* 151–159.

Carlyon, R. P. (1994). Detecting mistuning in the presence of synchronous and asynchronous interfering sounds. *Journal of the Acoustical Society of America, 95,* 2622–2630.

Cusack, R., & Roberts, B. (1999). Effects of similarity in bandwidth on the auditory sequential streaming of two-tone complexes. *Perception, 28,* 1281–1289.

Cusack, R., & Roberts, B. (2000). Effects of differences in timbre on sequential grouping. *Perception & Psychophysics, 62,* 1112–1120.

Cusack, R., & Roberts, B. (2004). Effects of differences in the pattern of amplitude envelopes across harmonics on auditory stream segregation. *Hearing Research, 193,* 95–104.

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. (1997). Perceptual segregation of a harmonic from a vowel by interaural time difference and frequency proximity. *Journal of the Acoustical Society of America, 102,* 2316–2324.

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., Hukin, R. W., & al-Khatib, B. Y. (1995). Grouping in pitch perception: Evidence for sequential constraints. *Journal of the Acoustical Society of America, 98,* 880–885.

Darwin, C. J., Pattison, H., & Gardner, R. B. (1989). Vowel quality changes produced by surrounding tone sequences. *Perception & Psychophysics, 45,* 333–342.

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

Dau, T., Ewert, S., & Oxenham, A. J. (2009). Auditory stream formation affects comodulation masking release retroactively. *Journal of the Acoustical Society of America, 125,* 2182–2188.

Deutsch, D. (1974). An auditory illusion. *Nature, 251,* 307–309.

Deutsch, D. (2004). The octave illusion revisited again. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 355–364.

Elhilali, M., Ma, L., Micheyl, C., Oxenham, A. J., & Shamma, S. A. (2009). Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron, 61,* 317–329.

Gockel, H., & Carlyon, R. P. (1998). Effects of ear of entry and perceived location of synchronous and asynchronous components on mistuning detection. *Journal of the Acoustical Society of America, 104,* 3534–3545.

Grimault, N., Bacon, S. P., & Micheyl, C. (2002). Auditory stream segregation on the basis of amplitude-modulation rate. *Journal of the Acoustical Society of America, 111,* 1340–1348.

Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., & Collet, L. (2000). Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency. *Journal of the Acoustical Society of America, 108,* 263–271.

Grose, J. H., & Hall, J. W. (1993). Comodulation masking release: Is

comodulation sufficient? *Journal of the Acoustical Society of America, 93,* 2896–2902.

Hafter, E. R., Bonnel, A. M., & Gallun, E. (1998). A role for memory in divided attention between two independent stimuli. In A. Palmer, A. Rees, A. Q. Summerfield, & R. Meddis (Eds.), *Psychophysical and physiological advances in hearing* (pp. 228–238). London: Whurr.

Hartmann, W. M., & Johnson, D. (1991). Stream segregation and peripheral channeling. *Music Perception, 9,* 155–184.

Hirsch, I. J. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America, 312,* 759–767.

Hong, R. S., & Turner, C. W. (2006). Pure-tone auditory stream segregation and speech perception in noise in cochlear implant recipients. *Journal of the Acoustical Society of America, 120,* 360–374.

Houtsma, A. J. M., & Goldstein, J. L. (1972). The central origin of the pitch of complex tones: Evidence from musical interval recognition. *Journal of the Acoustical Society of America, 51,* 520–529.

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.

Lee, A. K., Babcock, S., & Shinn-Cunningham, B. G. (2008). Measuring the perceived content of auditory objects using a matching paradigm. *Journal of the Association for Research in Otolaryngology, 9,* 388–397.

Lee, A. K., & Shinn-Cunningham, B. G. (2008). Effects of frequency disparities on trading of an ambiguous tone between two competing auditory objects. *Journal of the Acoustical Society of America, 123,* 4340–4351.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49,* 467–477.

Macmillan, N. A. (1971). Detection and recognition of increments and decrements in auditory intensity. *Perception & Psychophysics, 10,* 233–238.

Macmillan, N. A. (1973). Detection and recognition of intensity changes in tone and noise: The detection-recognition disparity. *Perception & Psychophysics, 13,* 65–75.

Micheyl, C., Kaernbach, C., & Demany, L. (2008). An evaluation of psychophysical models of auditory change perception. *Psychological Review, 115,* 1069–1083.

Miller, G., & Heise, G. A. (1950). The trill threshold. *Journal of the Acoustical Society of America, 22,* 637–638.

Moore, B. C. J., & Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acustica United with Acustica, 88,* 320–333.

Mossbridge, J. A., Fitzgerald, M. B., O'Connor, E. S., & Wright, B. A. (2006). Perceptual-learning evidence for separate processing of asynchrony and order tasks. *Journal of Neuroscience, 26,* 12708–12716.

O'Connor, K. N., & Sutter, M. L. (2000). Global spectral and location effects in auditory perceptual grouping. *Journal of Cognitive Neuroscience, 12,* 342–354.

Oertel, D., Bal, R., Gardner, S. M., Smith, P. H., & Joris, P. X. (2000). Detection of synchrony in the activity of auditory nerve fibers by octopus cells of the mammalian cochlear nucleus. *Proceedings of the National Academy of Sciences of the United States of America, 97,* 11773–11779.

Oxenham, A. J. (2000). Influence of spatial and temporal coding on auditory gap detection. *Journal of the Acoustical Society of America, 107,* 2215–2223.

Oxenham, A. J. (2008). Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants. *Trends in Amplification, 12,* 316–331.

Pastore, R. E., Harris, L. B., & Kaplan, J. K. (1982). Temporal order identification: Some parameter dependencies. *Journal of the Acoustical Society of America, 71,* 430–436.

Ringach, D. L., & Shapley, R. (1996). Spatial and temporal properties of illusory contours and amodal boundary completion. *Vision Research, 36,* 3037–3050.

Roberts, B., Glasberg, B. R., & Moore, B. C. (2002). Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. *Journal of the Acoustical Society of America, 112,* 2074–2085.

Roberts, B., Glasberg, B. R., & Moore, B. C. (2008). Effects of the build-up and resetting of auditory stream segregation on temporal discrimination. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 992–1006.

Rogers, W. L., & Bregman, A. S. (1998). Cumulation of the tendency to segregate auditory streams: Resetting by changes in location and loudness. *Perception & Psychophysics, 60,* 1216–1227.

Shinn-Cunningham, B. G., Lee, A. K., & Oxenham, A. J. (2007). A sound element gets lost in perceptual competition. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 12223–12227.

Sutter, M. L., Petkov, C., Baynes, K., & O'Connor, K. N. (2000). Auditory scene analysis in dyslexics. *Neuroreport, 11,* 1967–1971.

van Noorden, L. (1975). *Temporal coherence in the perception of tone sequences.* Unpublished doctoral dissertation, Eindhoven University of Technology, Eindhoven, the Netherlands.

van Noorden, L. P. (1977). Minimum differences of level and frequency for perceptual fission of tone sequences ABAB. *Journal of the Acoustical Society of America, 61,* 1041–1045.

Vliegen, J., Moore, B. C., & Oxenham, A. J. (1999). The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. *Journal of the Acoustical Society of America, 106,* 938–945.

Vliegen, J., & Oxenham, A. J. (1999). Sequential stream segregation in the absence of spectral cues. *Journal of the Acoustical Society of America, 105,* 339–346.

Warren, R. M. (2008). *Auditory perception: An analysis and synthesis.* Cambridge, United Kingdom: Cambridge University Press.

Warren, R. M., Obusek, C. J., Farmer, R. M., & Warren, R. P. (1969). Auditory sequence: Confusion of patterns other than speech or music. *Science, 164,* 586–587.

Wright, B. A., Buonomano, D. V., Mahncke, H. W., & Merzenich, M. M. (1997). Learning and generalization of auditory temporal-interval discrimination in humans. *Journal of Neuroscience, 17,* 3956–3963.

Zera, J., & Green, D. M. (1993a). Detecting temporal asynchrony with asynchronous standards. *Journal of the Acoustical Society of America, 93,* 1571–1579.

Zera, J., & Green, D. M. (1993b). Detecting temporal onset and offset asynchrony in multicomponent complexes. *Journal of the Acoustical Society of America, 93,* 1038–1052.

Zera, J., & Green, D. M. (1993c). Effect of signal component phase on asynchrony discrimination. *Journal of the Acoustical Society of America, 98,* 817–827.