Journal of Experimental Psychology:
Human Perception and Performance
2014, Vol. 40, No. 2, 685–701

0096-1523/14/$12.00 DOI: 10.1037/a0034720

Effects of Attention to and Awareness of Preceding Context Tones on Auditory Streaming

David M. Weintraub, Brian A. Metzger, and Joel S. Snyder
University of Nevada Las Vegas

This study determined whether facilitation of auditory stream segregation could occur when facilitating context tones are accompanied by other sounds. Facilitation was measured as the likelihood of a repeated context tone that could match the low (A) or high (B) frequency of a repeating ABA test to increase the likelihood of hearing the test as segregated. We observed this type of facilitation when matching tones were alone, or with simultaneous bandpass noises or continuous speech, neither of which masked the tones. However, participants showed no streaming facilitation when a harmonic complex masked the context tones. Mistuning or desynchronizing the context tone relative to the rest of the complex did not facilitate streaming, despite the fact that the context tone was accessible to awareness and attention. Even presenting the context tone in a separate ear from the rest of the harmonic complex did not facilitate streaming, ruling out peripheral interference. Presenting the test as mistuned or desynchronized tones relative to complex tones eliminated the possibility that timbre changes from context to test interfered with facilitation resulting from the context. These results demonstrate the fragility of streaming facilitation and show that awareness of and attention to the context tones are not sufficient to overcome interference.

Keywords: auditory scene analysis, awareness, attention, facilitation, frequency dependence

A central issue in the study of perception is the extent to which low-level factors and high-level factors contribute to detection, discrimination, categorization, and perceptual organization of stimuli. For example, several conceptual frameworks have been proposed about the nature of bottom-up information processing in the visual system and the extent to which attention and other top-down processes are required for normal perception (e.g., Bullier, 2001; Hochstein & Ahissar, 2002; Koch & Tsuchiya, 2007; Lamme, 2003; Treisman & Gelade, 1980). Over the past several decades, the contributions of low- and high-level factors to perception have been addressed extensively by studies of auditory scene analysis, the ability to segregate sounds from different sources into distinct auditory objects or streams (Bregman, 1990). Although some have made the argument that some types of auditory scene analysis tasks mainly rely on peripheral encoding of sound features to segregate sounds (Anstis & Saida, 1985; Beauvois & Meddis, 1996; Hartmann & Johnson, 1991; Van Noorden, 1975), most recent reviews of the literature have concluded that auditory scene analysis probably results from rich contributions of both low-level and high-level processes (Alain, 2007; Bregman, 1990; Moore & Gockel, 2002; Snyder & Alain, 2007; Snyder, Gregg, Weintraub, & Alain, 2012).

Much of the research on auditory scene analysis has examined auditory stream segregation, which is often studied by presenting alternating low-frequency tones (A), high-frequency tones (B), and silences (−) in a repeating ABA− pattern. Perception of two segregated streams is promoted when the frequency separation (Δf) between the A and B tones is large and when the presentation rate of the tones is fast (Bregman & Campbell, 1971; Van Noorden, 1975). Additionally, perception of one stream is usually heard at the beginning of a repeating ABA− sequence but the likelihood of perceiving two streams (or "streaming") increases as the number of ABA− repetitions increases (Anstis & Saida, 1985; Bregman, 1978), a phenomenon called 'build-up.' In addition, streaming can be facilitated by prior presentation of several repetitions of tones with the same frequency as one of the tones of the following ABA− pattern (Beauvois & Meddis, 1997; Rogers & Bregman, 1993, 1998), which may result from at least partially distinct mechanisms compared with buildup (Haywood & Roberts, in press).

Several forms of evidence suggest that stream segregation involves high-level processes. For example, although stream segregation is most robust for cues extracted in the periphery (i.e., frequency and ear of presentation, Hartmann & Johnson, 1991), cues extracted in the central auditory system can also lead to streaming (Cusack & Roberts, 2000; Grimault, Bacon, & Micheyl, 2002; Vliegen, Moore, & Oxenham, 1999; Vliegen & Oxenham, 1999). Similarly, the effect of prior Δf on perception of streaming

This article was published Online First November 4, 2013.

David M. Weintraub, Brian A. Metzger, and Joel S. Snyder, Department of Psychology, University of Nevada Las Vegas.

Brian A. Metzger is now in the Department of Psychology, University of Illinois at Urbana-Champaign.

We thank Dr. James Enns and three anonymous reviewers whose feedback greatly improved this paper. We were supported by National Science Foundation Grant BCS1026023 (J.S.S. and D.M.W.) and a summer research stipend from the College of Liberal Arts at the University of Nevada Las Vegas (J.S.S.).

Correspondence concerning this article should be addressed to Joel S. Snyder, Department of Psychology, University of Nevada Las Vegas, 4505 South Maryland Parkway Box 455030, Las Vegas, NV, 89154-5030. E-mail: Joel.Snyder@unlv.edu

(Snyder, Carter, Lee, Hannon, & Alain, 2008) is likely to occur at higher levels of the auditory system sensitive to rhythmic patterns, rather than at early tonotopic levels (Snyder, Carter, Hannon, & Alain, 2009; Snyder, Holder, Weintraub, Carter, & Alain, 2009; Snyder & Weintraub, 2011), and to arise from memories that also contain information about prior perceptual interpretations (Snyder & Weintraub, 2013).

In addition to these lines of evidence, several studies have addressed whether attention, which we define as the selective processing of certain stimuli (e.g., ABA− sounds, as opposed to other co-occurring auditory or visual stimuli), is *necessary* for streaming to occur. On the one hand, neural correlates of streaming have been observed in the cochlear nucleus of anesthetized animals (Pressnitzer, Sayles, Micheyl, & Winter, 2008), similar to activity observed in primary auditory cortex in animals that were not making perceptual decisions (Micheyl, Tian, Carlyon, & Rauschecker, 2005). Similarly, in humans, a mismatch negativity event-related potential dependent on segregation of streams was found to increase in amplitude later in a tone sequence as would be expected by the time course of streaming, even though participants ignored the sounds (Sussman, Horvath, Winkler, & Orr, 2007). These studies showing neural correlates of streaming in the absence of perceptual judgments suggest that attention is not necessary for streaming and, instead, streaming may rely to a large extent on low-level factors. On the other hand, a number of neurophysiological and behavioral studies directly manipulating attention have come to the opposite conclusion, namely that streaming is strongly diminished by lack of attention to the tone patterns or by the attention switching that sometimes is required in these paradigms (Carlyon, Cusack, Foxton, & Robertson, 2001; Carlyon, Plack, Fantini, & Cusack, 2003; Cusack, Deeks, Aikman, & Carlyon, 2004; Elhilali, Xiang, Shamma, & Simon, 2009; Snyder, Alain, & Picton, 2006; Thompson, Carlyon, & Cusack, 2011).

In contrast to these previous studies, the current study asks whether attention to and awareness of a context tone sequence are *sufficient* to facilitate streaming, even in the presence of interfering sounds. In other words, is conscious perception of a context tone attributable to it being processed in the peripheral and the central auditory system enough for it to facilitate streaming? Or can facilitation be interfered with by the presence of other concurrent sounds during the context? We studied streaming facilitation using a paradigm inspired by previous studies in which a repeating context tone of a fixed frequency was presented that matched one of the tones in a following ABA− pattern (see Figure 1, Beauvois & Meddis, 1997; Haywood & Roberts, in press; Rogers & Bregman, 1993, 1998). To manipulate attention to and awareness of the repeating context tone, we embedded it among other simultaneous tones, which are known to prevent it from standing out as a separate auditory object (Elhilali, Ma, Micheyl, Oxenham, & Shamma, 2009). Thus, we were able to determine whether a context tone can facilitate streaming even when participants are not aware of the tone's presence (i.e., while not reporting hearing the tone) and thus in all likelihood also not selectively attending to it before the presentation of an ABA− pattern (Experiment 1). Next, by mistuning the context tone or desynchronizing it relative to the rest of the simultaneous tones in the context, and thus making it stand out as a separate auditory object (Moore, Peters, & Glasberg, 1985), we were able to study whether attention to and awareness of the context tone are sufficient to facilitate streaming (Experiments 2–6). Surprisingly, we demonstrate that in the presence of other tones, attention to and awareness of the critical context tone is not sufficient to facilitate streaming. This is likely because of some form of high-level interference between the critical context tone and other context tones that have similar timbre and temporal patterning, as suggested by the final experiment we report (Experiment 7).

Experiment 1

Method

Participants. Twelve adults with normal self-reported hearing (six men and six women, age range = 18–40 years, mean age = 22.75 years) from the University of Nevada, Las Vegas Psychology subject pool participated after giving written informed consent according to the guidelines of the University's Office for the Protection of Research Subjects.

Stimuli and procedure. The stimuli were tones 50 ms in duration, including 5 ms rise/fall times with linear ramps. The stimuli were synthesized off-line using Matlab (The MathWorks Inc., Natick, MA) and presented using a custom interface written in Presentation (Neurobehavioral Systems, Inc., Albany, CA). Sounds were generated using an SB X-Fi sound card (Creative Technology, Ltd.) and delivered binaurally via Sennheiser HD 280 headphones (Sennheiser Electronic Corporation, Old Lyme, CT) at approximately 70 dB SPL.

Participants were seated in a quiet room and were asked to maintain fixation on a white cross on a black background in the center of a computer screen throughout the experiment, to minimize potential visual influences on the auditory percepts. Each trial consisted of two successive tone sequences: a variable "context" consisting of a sequence of simple or complex tones, and a constant "test" consisting of a sequence of simple tones (see Figure 1). The test and context sequences were separated by a 120 ms stimulus onset asynchrony (SOA) between the final tone of the context and the first tone of the test, which resulted in a seamless presentation with no disruption of rhythm. Consecutive trials were separated by an interval of 1.94 s or 5.00 s intermixed randomly between the end of the final tone of the test of the previous trial and the first tone of the context sequence of the next trial. The test

Figure 1 (opposite). Experimental stimuli. The top of the figure shows the general trial structure of all experiments, with a context sequence composed of 28 tones with an SOA of 240 ms from one tone to the next, and a test sequence composed of 14 ABA− tone patterns with an SOA of 480 ms from one ABA− to the next. Below, the context and test tones for each experiment are shown. Solid lines refer to simple tones or tones that are tuned (i.e., integer multiple) with respect to the fundamental of a complex tone. Dotted lines refer to mistuned (i.e., not integer multiple) tones. For Experiment 3, c and i stand for contralateral and ipsilateral with respect to the test tones.

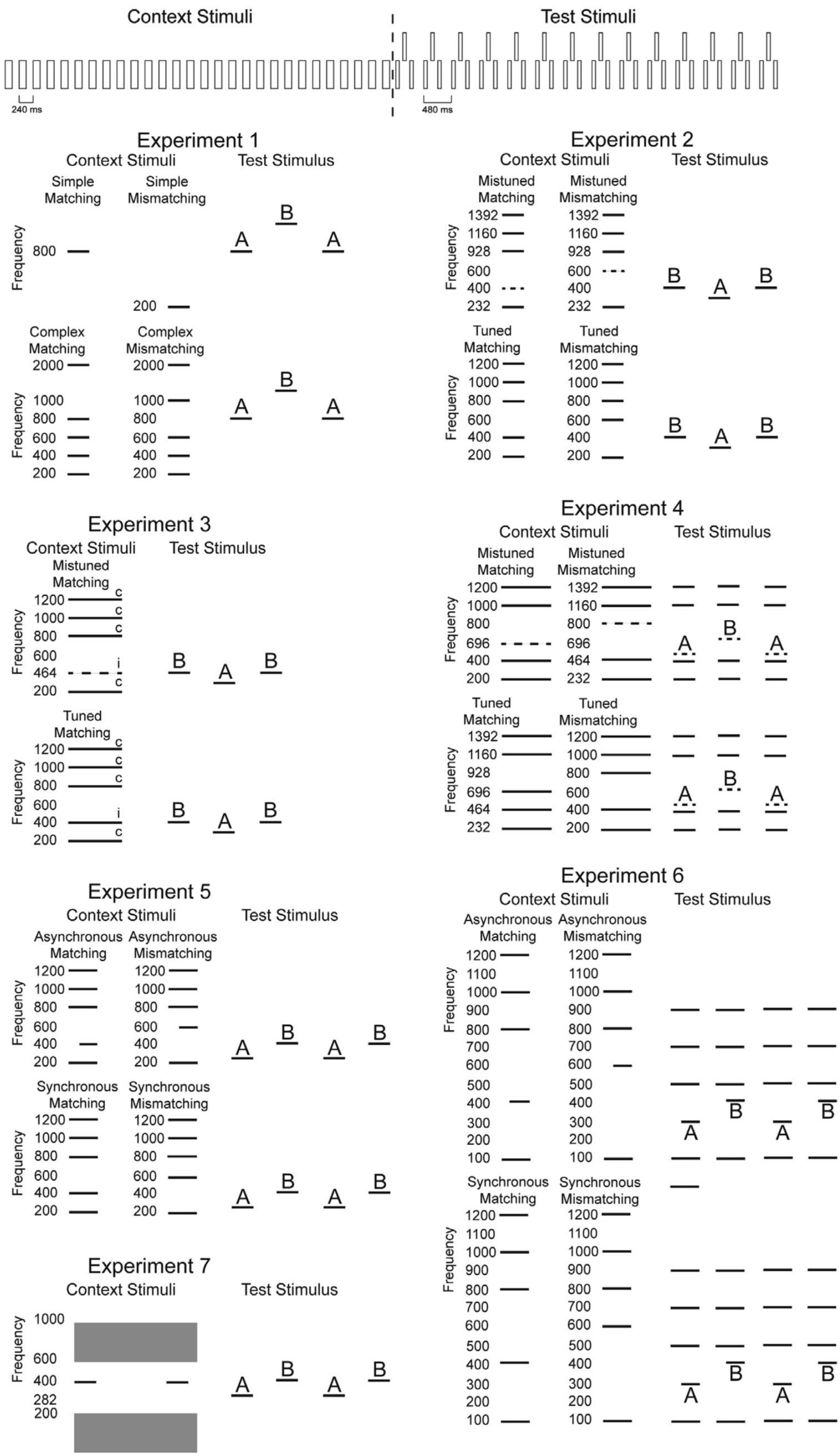

Figure 1 (opposite)

sequence was always the same across conditions and consisted of 14 ABA− triplets, in which consecutive A and B tones within a triplet had an SOA of 120 ms and the silent duration (−) also was 120 ms. This resulted in an SOA between consecutive ABA− triplets of 480 ms, for a total test sequence duration of 6.72 s. The frequency of the A tones was fixed at 800 Hz and the frequency of the B tones was fixed at 1270 Hz. This corresponds to an A-B Δf of 8.0 semitones. This Δf was chosen because it usually leads to an ambiguous, or bistable, percept, meaning, the sequence can be perceived as a single stream, or as two separate streams (Van Noorden, 1975).

There were four types of context sequences, two containing simple (i.e., pure) tones and two containing complex tones. Of the two simple and two complex context sequences, one sequence contained a tone matching the frequency of the A tone in the test sequence and the other sequence contained no tones matching the frequency of the A tone in the test sequence. All context sequences consisted of 28 tones (except those containing an omission, explained below), each separated from the next by an SOA of 240 ms, for a context sequence duration of 6.72 s. One of the simple conditions, called the *simple-matching* condition, had simple tones that matched the 800 Hz frequency of the A tone in the test sequence. The other simple condition, called the *simple-mismatching* condition, had simple tones that did not match the A tone in the test sequence, instead being presented at 200 Hz. One of the complex conditions, called the *complex-matching* condition, had complex harmonic tones with a fundamental frequency of 200 Hz, and the following integer-multiple harmonics: 400, 600, 800, and 2000 Hz. Thus, in this condition, the fourth harmonic (i.e., 800 Hz) matched the frequency of the A tone of the test sequence. The other complex condition, called the *complex-mismatching* condition, had complex harmonic tones with a fundamental frequency of 200 Hz, and the following integer-multiple harmonics: 400, 600, 1000, and 2000 Hz, importantly not including the 800 Hz harmonic that corresponds to the frequency of the A tone in the test sequence. On 12.69% of the trials, one of the context tones (randomly chosen from any tone between and including the 5th and 10th tones) was omitted to make sure participants were attending to the context. In addition to the four trial types with context sequences, an additional trial type was included in which the context period was filled with 6.72 s of silence, which served as a baseline condition.

Two blocks of 31 trials and two blocks of 32 trials were presented. This allowed us to present 126 trials in total including 22 of each of the five trial types and 16 trials containing omissions. The different types of trials were randomly intermingled within a block. Before the experiment, we presented each of the five trial types two times as practice in random order. During the context, participants were instructed to listen carefully to the sequence and press the ‘0’ button on the keyboard (number pad) whenever a tone omission was detected. During the test, participants were instructed to press and hold down the key labeled ‘1’ on the computer keyboard whenever they perceived the sound sequence as a single stream, and to press and hold down the key labeled ‘2’ whenever they perceived the sound sequence as two streams. They were asked to release all keys during silent intervals. In addition, the participants were encouraged to not actively try to hear the sequence one way or the other but, rather, to listen “neutrally” and attentively to the sequences. Button presses and releases were recorded by Presentation, and stored for off-line analysis.

Data analysis. Key presses within 1.92 s after an omission during the context were counted as successful detections. Test sequences that followed a context sequence containing an omission were not included for analysis. The timings of the key presses during the test were used to construct the time courses of perceiving “two streams,” separately, for each experiment, trial type, and participant. The time series for each test represented a total duration of 6.72 s and consisted of 14 time points. Each time point represented the instantaneous reported perception during a time span beginning at an ABA− cycle (i.e., every 480 ms). Each of the time points was coded as having no response if no button had been pressed previously during the trial. A data point was coded as “1-stream” only if the “1-stream” button had been pressed most recently during the trial, or if the “1-stream” button was pressed immediately after the current time point and closer to the current time point than the next time point. A data point was coded as “2 streams” only if the last previous button pressed during the trial was the “2-streams” button, or if the “2-streams” button was pressed immediately after the current time point and closer to the current time point than the next time point. For each participant, we calculated the proportion of trials during which participants reported perceiving two streams for each time point within a condition by averaging the time series across all the trials within the same condition; these average time courses are plotted in the figures. To quantify streaming for statistical analysis, we calculated the proportion of total time that each participant reported two streams by averaging the last 11 time points together from the test period. The first three time points from every condition were excluded from the analyses to avoid including time points without button presses.

The proportion of omissions detected during the context and the proportions of hearing two streams during the test for the four different trial types with nonsilent contexts were entered into separate two-factor repeated-measures analyses of variance (ANOVAs) to test for differences depending on whether the context had simple or complex tones and whether it did or did not contain the matching 800 Hz tone. An additional one-factor repeated-measures ANOVA on the test data was performed comparing the silent condition to the three other conditions not including the simple-matching condition. p values less than .05 were considered statistically significant, and when appropriate Greenhouse-Geisser corrected p values were reported.

Results and Discussion

Participants were generally successful in performing the omission detection task during the context. On average, participants detected 79.49% of omissions. Additionally, performance was not different across all trial types, $F(3, 33) = 0.55$, $p = .574$, $\eta_p^2 = .047$. Performance was not affected by the tone type (simple vs. complex), $F(1, 11) = 0.34$, $p = .571$, $\eta_p^2 = .030$, or whether the context was matching or mismatching the test, $F(1, 11) = 0.51$, $p = .489$, $\eta_p^2 = .045$, and there was no interaction between these two factors, $F(1, 11) = 0.75$, $p = .406$, $\eta_p^2 = .064$.

Figure 2 shows the time course of the “two streams” percept, averaged across all participants during Experiment 1. In general, participants followed instructions by continuously indicating

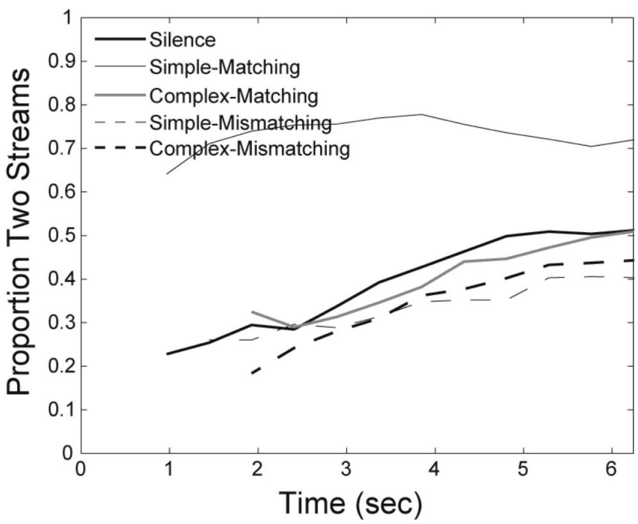

Figure 2. Time course data of streaming during the test from Experiment 1. Only the condition with a simple-tone matching context facilitated streaming of the ABA− test pattern.

whether they were hearing one or two streams during the test periods, and ceasing to press buttons during the silent periods. As shown in Figure 2, effects that have been observed in numerous previous studies were replicated (e.g., Anstis & Saida, 1985; Bregman, 1978; Carlyon et al., 2001). In particular, at the beginning of the sequences, participants tended to perceive one stream and only after multiple repetitions of the ABA− pattern did perception tend to switch to two streams. Furthermore, the context that consisted of a repeating, attended tone that matched the A tone of the test (simple-matching condition) greatly facilitated perception of two streams during the test, compared with a silent context and compared with a context that consisted of a repeating tone that did not match either the A or B tone of the test (simple-mismatching condition).

The more novel conditions containing complex tones (complex-matching and complex-mismatching) did not facilitate perception of streaming compared with the silent or simple-mismatching conditions, even when the context tones contained the A tone as a harmonic (complex-matching). This pattern of results with only one of the conditions (simple-matching) being different than the rest was substantiated by the two-factor ANOVA, in which there were significant main effects of tone type, $F(1, 11) = 9.21$, $p < .025$, $\eta_p^2 = .456$, and matching, $F(1, 11) = 20.10$, $p < .001$, $\eta_p^2 = .646$, and a significant interaction between these two factors, $F(1, 11) = 10.01$, $p < .01$, $\eta_p^2 = .476$. The fact that the conditions other than the simple-matching condition were no different than silence was confirmed by the one-factor ANOVA, in which the condition factor was not significant, $F(3, 33) = 1.25$, $p = .30$, $\eta_p^2 = .102$. In fact, the trend was for the nonsilent conditions to show a bit less streaming than the silent condition.

The results of this experiment were successful in demonstrating the frequency-dependence of streaming when only a simple tone (that matched the A tone of the test ABA− pattern) was presented during the context to facilitate perception of two streams during the test (cf. Anstis & Saida, 1985; Roberts et al., 2008). However, the results failed to demonstrate that the same type of frequency-dependent streaming could occur when the same matching tone was masked from awareness by presenting it as the fourth harmonic of a complex tone. Although this negative finding could be explained by a number of factors, the results clearly show that a tone that is likely to be resolved in the periphery (Glasberg & Moore, 1990) and thus cause adaptation of frequency-selective neurons in early stages of auditory processing is not sufficient to facilitate streaming. This clearly violates any strong form of the peripheral channeling model of streaming (Beauvois & Meddis, 1996; Hartmann & Johnson, 1991), which posits that all basic aspects of streaming can be explained by frequency-dependent processing in the ear and auditory nerve. Similarly, observations of frequency-specific neural adaptation that has been observed in the brainstem and auditory cortex may reflect processing that is necessary but not sufficient for streaming to occur (Micheyl et al., 2005; Pressnitzer et al., 2008).

As mentioned above, there are several factors that may be necessary for facilitation of streaming to occur in addition to frequency-specific adaptation that were prevented by embedding the critical tone as the fourth harmonic in a complex tone. First, it is possible that the role the critical tone played as a component of the complex tone object removed the critical tone's eligibility to influence perception of other objects, namely the stream of A tones. This is consistent with the idea of 'exclusive allocation,' promoted by Bregman (1990), which proposes that an individual sound component does not generally contribute to the perception of multiple objects (for violations of this principle, see Fowler & Rosenblum, 1990; Shinn-Cunningham, Lee, & Oxenham, 2007; reviewed by Snyder et al., 2012). A second, related possibility is that the inability to consciously hear the critical context tone prevented it from influencing later perception of streaming. Finally, and also related to the other factors, is the possibility that the inability to attend to the critical tone prevented it from facilitating streaming during the test. This latter possibility is in line with the findings suggesting that paying attention to A and B sounds may be important for streaming to occur (e.g., Carlyon et al., 2001; Snyder et al., 2006; Thompson et al., 2011; but see Sussman et al., 2007; reviewed by Snyder et al., 2012). Thus, although it is not possible to clearly identify the extent to which these or other factors are required for the facilitation of streaming to occur, it is clear that one or more high-level influences are required. The remaining experiments are meant to shed light on the necessary and sufficient conditions for streaming to occur, above and beyond frequency-dependent adaptation.

Experiments 2–6

This group of experiments was meant to test whether making the critical context tone stand out as a separate auditory object is sufficient for it to facilitate streaming. We manipulated these factors during the context in a variety of ways: mistuning the critical tone relative to its expected frequency as an integer-multiple harmonic of the fundamental frequency of a complex tone (Moore et al., 1985); increasing the duration of the context tones (cf. Alain, Schuler, & McDonald, 2002; Moore et al., 1985); delaying the onset of the critical tone relative to the rest of the context sounds (Elhilali, Ma et al., 2009); and by presenting the critical tone to the opposite ear as the rest of the context sounds. We also tried presenting the test sequences as mistuned harmonics in larger complexes to make the context and test more acoustically similar to each other (Anstis & Saida, 1985; Cusack et al., 2004; Roberts et al., 2008; Rogers & Bregman, 1993, 1998), which might prevent any resetting of streaming facilitation that occurs during the context. During the context, we asked participants to report whether they heard the critical tone popping out apart from the complex tone, before making streaming judgments about test sequences. Tables 1 and 2 summarize the important methodological parameters for the context and test sequences, respectively, for Experiments 2–6, which are described in more detail in the Method section.

Method

Participants. All experiments recruited participants with normal self-reported hearing from the University of Nevada, Las Vegas Psychology subject pool, who participated after giving written informed consent according to the guidelines of the University's Office for the Protection of Research Subjects. For Experiment 2, 10 adults participated (six men and four women, age range = 18–24 years, mean age = 19.4 years). For Experiment 3, 14 adults participated (six men and 10 women, age range = 18–38 years, mean age = 20.93 years). For Experiment 4, 18 adults participated (10 men and eight women, age range = 18–24 years, mean age = 20.39 years). For Experiment 5, 11 adults participated (seven men and four women, age range = 18–44 years, mean age = 22.45 years). For Experiment 6, 22 adults participated (eight men and 14 women, age range = 18–31 years, mean age = 19.91 years).

Stimuli and procedure. In Experiment 2, we mistuned the critical tone relative to its expected frequency as an integer-multiple harmonic of the fundamental frequency of a complex tone to make it perceptually salient (Moore et al., 1985). The stimuli and procedures for Experiment 2 were the same as in Experiment 1, except as follows. The context and test tones were 80 ms in duration, including 10-ms rise/fall times with linear ramps. Each test sequence consisted of 14 BAB− triplets. As shown in Figure 1, the frequency of the A tones was fixed at 283 Hz and the frequency of the B tones was fixed at 400 Hz. This corresponds to an A-B Δf of 6.00 semitones. There were four context sequences, all consisting of complex tones: 1) the *mistuned-matching* condition consisted of complex tones with a fundamental frequency of 232 Hz, a second harmonic of 400 Hz (mistuned down from 464 Hz) that matched the B tone frequency of the test, and integer-multiple harmonics of 928, 1160, and 1392 Hz; 2) the *mistuned-mismatching* condition consisted of complex tones with a fundamental frequency of 232 Hz, a missing second harmonic, a mistuned third harmonic of 600 Hz (mistuned down from 696 Hz), and integer-multiple harmonics of 928, 1160, and 1392 Hz; 3) the *tuned-matching* condition consisted of complex tones with a fundamental frequency of 200 Hz, a tuned second harmonic of 400 Hz that matched the B tone frequency of the test, and integer-multiple harmonics of 800, 1000, and 1200 Hz; and 4) the *tuned-mismatching* condition consisted of complex tones with a fundamental frequency of 200 Hz, a missing second harmonic, and integer-multiple harmonics of 600, 800, 1000, and 1200 Hz. As in Experiment 1, an additional trial type was included in which the context period was filled with 6.72 s of silence, which served as a baseline condition. Unlike in Experiment 1, there were no omitted tones during the context. Instead, we asked participants to report whether they were hearing the mistuning by continuously pressing

Table 1
Context Stimulus Parameters for Experiments 2–6

Experiment	Context tone type	Context tone frequencies[a]	Context tone duration	Context tone SOA	Context-test SOA
2	Complex	*Mistuned-Matching*: 232, **400**, 928, 1160, 1392 Hz *Mistuned-Mismatching*: 232, 600, 928, 1160, 1392 Hz *Tuned-Matching*: 200, **400**, 800, 1000, 1200 Hz *Tuned-Mismatching*: 200, 600, 800, 1000, 1200 Hz	80 ms	240 ms	120 ms
3	Dichotic, Complex	*Mistuned-Matching Left Ear:* Left: **464**; Right: 200, 800, 1000, 1200 Hz *Mistuned-Matching Right Ear*: Left: 200, 800, 1000, 1200 Hz; Right: **464** Hz *Tuned-Matching Left Ear*: Left: **400**; Right: 200, 800, 1000, 1200 Hz *Tuned-Matching Right Ear*: Left: 200, 800, 1000, 1200 Hz; Right: **400** Hz	200 ms	240 ms	240 ms
4	Complex	*Mistuned-Matching*: 200, 400, **696**, 1000, 1200 Hz *Mistuned-Mismatching*: 232, 464, 800, 1160, 1392 Hz *Tuned-Matching*: 232, 464, **696**, 1160, 1392 Hz *Tuned-Mismatching*: 200, 400, 800, 1000, 1200 Hz	200 ms	240 ms	240 ms
5	Asynchronous, Complex	*Asynchronous-Matching*[b]: 200, **400**, 800, 1000, 1200 Hz *Asynchronous-Mismatching*[b]: 200, 600, 800, 1000, 1200 Hz *Synchronous-Matching*: 200, **400**, 800, 1000, 1200 Hz *Synchronous-Mismatching*: 200, 600, 800, 1000, 1200 Hz	100 ms	120 ms	1.56 s
6	Asynchronous, Complex	*Asynchronous-Matching*[b]: 100, **400**, 800, 1000, 1200 Hz *Asynchronous-Mismatching*[b]: 100, 600, 800, 1000, 1200 Hz *Synchronous-Matching*: 100, **400**, 800, 1000, 1200 Hz *Synchronous-Mismatching*: 100, 600, 800, 1000, 1200 Hz	100 ms	120 ms	1.56 s

[a] Bolded tones correspond to critical context tones which preceded matching test tones. [b] Underlined tones correspond to asynchronous tones delayed by 40 ms relative to the onset of the remaining complex tones.

Table 2
Test Stimulus Parameters for Experiments 2–6

Experiment	Test tone type	Test tone frequencies[a]	Test tone duration	Test tone SOA	Test tone pattern
2	Simple	A = 283 Hz; **B = 400 Hz**	80 ms	120 ms	BAB
3	Simple	*Mistuned-Matching Left Ear*: Left: A = 315 Hz, **B = 464 Hz**; Right: silent *Mistuned-Matching Right Ear*: Left: silent; Right: A = 315 Hz, **B = 464 Hz** *Tuned-Matching Left Ear*: Left: A = 272 Hz, **B = 400 Hz**; Right: silent *Tuned-Mismatching Right Ear*: Left: silent; Right: A = 272 Hz, **B = 400 Hz**	80 ms	120 ms	BAB
4	Complex	A = 504 Hz + 200, 400, 1000, 1200 Hz; **B = 696 Hz** + 200, 400, 1000, 1200 Hz	80 ms	120 ms	ABA
5	Simple	A = 283 Hz; **B = 400 Hz**	80 ms	120 ms	ABAB
6	Complex	A = <u>300 Hz</u>[b] + 100, 500, 700, 900 Hz; **B** = **<u>400 Hz</u>**[b] + 100, 500, 700, 900 Hz	100 ms	120 ms	ABAB

[a] Bolded tones correspond to critical test tones preceded by matching tones in matching context conditions. [b] Underlined tones correspond to asynchronous tones delayed by 40 ms relative to the onset of the remaining complex tones.

the ‘0’ button on the keyboard (number pad) when they heard a simple “beeping tone” in addition to a complex “buzzing tone.” Participants were instructed to not push any button when they only heard the buzzing tone. Two blocks of 31 trials and two blocks of 32 trials were presented. Each nonsilent trial type was presented 26 times, and the silent trial was presented 22 times. The different types of trials were randomly intermingled within a block. Before the experiment, we presented six practice trials (each nonsilent context trial type was played once and the silent context trial type was played twice) in random order.

Despite the unlikely possibility that peripheral interference between the critical tone and the rest of the complex tone in the context prevented facilitation of streaming, we conducted Experiment 3 to rule this out more definitively. We did this by presenting the critical tone of the context in just one ear, and the rest of the complex in the opposite ear, followed by the ABA− test sequence in the same ear as the critical tone. We also lengthened the duration of the context tones, to further facilitate pop-out of the mistuned harmonic (cf. Alain, Schuler, & McDonald, 2002; Moore et al., 1985). The stimuli and procedure for Experiment 3 were the same as in Experiment 2, except as follows. In the mistuned condition, the frequency of the A tones of the test sequence was fixed at 315 Hz and the frequency of the B tones was fixed at 464 Hz. This corresponds to an A-B Δf of 6.71 semitones. In the tuned condition, the frequency of the A tones of the test sequence was fixed at 272 Hz and the frequency of the B tones was fixed at 400 Hz. This corresponds to an A-B Δf of 6.68 semitones. Different test sequences were used for these two conditions so that their B tone always matched the frequency of second harmonic in the preceding context sequence (as discussed below). On each trial, the test tones were all presented either to the left ear or the right ear. The context tones were 200 ms in duration, including 10 ms rise/fall times with linear ramps. There were two context sequences, both consisting of complex tones: 1) the *mistuned-matching* condition consisted of complex tones with a fundamental frequency of 200 Hz, a second harmonic of 464 Hz (mistuned up from 400 Hz) that matched the B tone frequency of the test, and integer-multiple harmonics of 800, 1000, and 1200 Hz; and 2) the *tuned-matching* condition consisted of complex tones with a fundamental frequency of 200 Hz, a second harmonic at 400 Hz that matched the B tone frequency of the test, and integer-multiple harmonics of 800, 1000, and 1200 Hz. Importantly, the second harmonic of the context tones was always presented to the same (ipsilateral) ear as the test tones; all of the other harmonics were always presented to the opposite (contralateral) ear as the test tones. The second harmonic of the context was either presented in the left or right ear. As in the previous experiments, an additional trial type was included in which the context period was filled with 6.72 s of silence, which served as a baseline condition. Thus, there were a total of six trial types depending on the context type and the ear of presentation for the test tones: *mistuned-matching left ear*, *mistuned-matching right ear*, *tuned-matching left ear*, *tuned-matching right ear*, *silence left ear*, and *silence right ear*. Four blocks of 48 trials were presented (32 of each of the six trial types). The different types of trials were randomly intermingled within a block. Before this experiment, we presented five practice trials such that all nonsilent context trial types and the *silence right ear* trial type were each played one time.

In Experiment 4, we considered whether a change from presenting complex tones during the context to simple tones during the test might have reset facilitation that may have indeed occurred during the matching context conditions, especially those in which the critical tone was mistuned and therefore consciously accessible to perception. Such resetting might be expected based on several previous studies showing that a variety of acoustic and attentional changes can reset the streaming process (Anstis & Saida, 1985; Cusack et al., 2004; Kondo, Pressnitzer, Toshima, & Kashino, 2012; Roberts et al., 2008; Rogers & Bregman, 1993, 1998). To test this possibility, we used similar complex-tone context sequences as in the previous experiments, but unlike the previous experiments also presented the A and B test tones as mistuned harmonics. The stimuli and procedure for Experiment 4 were the same as in Experiment 2, except as follows. The test tones were 80 ms in duration, including 10-ms rise/fall times with linear ramps. As shown in Figure 1, the frequency of the A tones was fixed at 504 Hz and the frequency of the B tones was fixed at 696 Hz. This corresponds to an A-B Δf of 5.59 semitones. Unlike in previous experiments, however, the A and B tones of the test sequence were presented as mistuned harmonics of complex tones. Thus, the following additional tones were presented simultaneously with both the A and B tones: 200, 400, 1000, and 1200 Hz. This means that the A tone was mistuned down from 600 Hz, whereas the B tone was mistuned up from 600 Hz. The context tones were 200 ms in duration, including 10-ms rise/fall times with linear ramps.

There were four context sequences, all consisting of complex tones: 1) the *mistuned-matching* condition consisted of complex tones with a fundamental frequency of 200 Hz, a second harmonic of 400 Hz, a third harmonic of 696 Hz (mistuned up from 600 Hz) that matched the B tone frequency of the test, and integer-multiple harmonics of 1000 and 1200 Hz; 2) the *mistuned-mismatching* condition consisted of complex tones with a fundamental frequency of 232 Hz, a second harmonic of 464 Hz, a missing third harmonic, a fourth harmonic of 800 Hz (mistuned down from 928 Hz), and integer-multiple harmonics of 1160 and 1392 Hz; 3) the *tuned-matching* condition consisted of complex tones with a fundamental frequency of 232 Hz, tuned second and third harmonics of 464 and 696 Hz, and integer-multiple harmonics of 1160 and 1392 Hz; and 4) the *tuned-mismatching* condition consisted of complex tones with a fundamental frequency of 200 Hz, a second harmonic of 400 Hz, a missing third harmonic, and additional integer-multiple harmonics of 800, 1000, and 1200 Hz. An additional trial type was included in which the context period was filled with 6.72 s of silence, which served as a baseline condition. Four blocks of 40 trials were presented (32 of each of the five trial types). The different types of trials were randomly intermingled within a block. Before the experiment, we presented each of the five trials types two times as practice in random order.

The purpose of Experiment 5 is to determine the effect of using a temporal asynchrony cue (instead of a mistuning cue) to separate the critical tone from the other context tones on streaming during the test. The stimuli and procedure for Experiment 5 were the same as in Experiment 2, except as follows. The test stimuli were presented in an ABAB pattern (i.e., in an isochronous fashion) rather than an ABA− pattern. The test tones were 80 ms in duration, including 10-ms rise/fall times with linear ramps. As shown in Figure 1, the frequency of the A tones was fixed at 283 Hz and the frequency of the B tones was fixed at 400 Hz. This corresponds to an A-B Δf of 6 semitones. Except for the asynchronous critical tones, the context tones were 100 ms in duration, including 10-ms rise/fall times with linear ramps. There were four context sequences, all consisting of complex tones: 1) the *asynchronous-matching* condition consisted of complex tones with a fundamental frequency of 200 Hz, a second harmonic of 400 Hz that matched the B tone frequency of the test but had an onset that was 40 ms later than the other tones (but with common offset time), and integer-multiple harmonics of 800, 1000, and 1200 Hz; 2) the *asynchronous-mismatching* condition consisted of complex tones with a fundamental frequency of 200 Hz, an additional harmonic of 600 Hz that had an onset that was 40 ms later than the other tones (but with common offset time), and integer-multiple harmonics of 800, 1000, and 1200 Hz; 3) the *synchronous-matching* condition consisted of complex tones with a fundamental frequency of 200 Hz, and harmonics of 400, 800, 1000, and 1200 Hz that were all synchronous with the fundamental; and 4) the *synchronous-mismatching* condition consisted of complex tones with a fundamental frequency of 200 Hz, and harmonics of 600, 800, 1000, and 1200 Hz that were all synchronous with the fundamental. An additional trial type was included in which the context period was filled with 6.72 s of silence, which served as a baseline condition. Unlike the previous experiments, there was a 1.44-s silent period between the context and test sequences for all conditions. It is unlikely that any null results reported in this experiment are attributable to the short silent period given that previous studies have shown the effects of streaming facilitation are strong even with a 1.44 s gap and do not fully decay until about 4 s (Beauvois & Meddis, 1997). Five blocks of 30 trials were presented (30 of each of the five trial types). The different types of trials were randomly intermingled within a block. Before the experiment, we presented 8 practice trials chosen randomly from all 5 trial types with the exception that no trial type was played more than twice.

The stimuli and procedure for Experiment 6 were the same as in Experiment 5, except as follows. The test tones were 100 ms in duration, including 10-ms rise/fall times with linear ramps. As shown in Figure 1, the frequency of the A tones was fixed at 300 Hz and the frequency of the B tones was fixed at 400 Hz. This corresponds to an A-B Δf of 5 semitones. A and B tones of the test sequence were presented as asynchronous harmonics of complex tones. Thus, the following additional tones were presented simultaneously with both the A and B tones: 100, 500, 700, and 900 Hz. A and B tones were presented with an onset 40 ms after the onset of the other tones. All harmonics in the same complex tone had a common offset time. The context tones were identical to those in Experiment 5 except as follows. There were four context sequences, all consisting of complex tones: 1) the *asynchronous-matching* condition consisted of complex tones with a fundamental frequency of 100 Hz, an additional harmonic of 400 Hz that matched the B tone frequency of the test but had an onset that was 40 ms later than the other tones (but with common offset time), and integer-multiple harmonics of 800, 1000, and 1200 Hz; 2) the *asynchronous-mismatching* condition consisted of complex tones with a fundamental frequency of 100 Hz, an additional harmonic of 600 Hz that had an onset that was 40 ms later than the other tones (but with common offset time), and integer-multiple harmonics of 800, 1000, and 1200; 3) the *synchronous-matching* condition consisted of complex tones with a fundamental frequency of 100 Hz, and harmonics of 400, 800, 1000, and 1200 Hz that were all synchronous with the fundamental; and 4) the *synchronous-mismatching* condition consisted of complex tones with a fundamental frequency of 100 Hz, and harmonics of 600, 800, 1000, and 1200 Hz that were all synchronous with the fundamental. An additional trial type was included in which the context period was filled with 6.72 s of silence, which served as a baseline condition. Five blocks of 30 trials were presented (30 of each of the five trial types). The different types of trials were randomly intermingled within a block. Before the experiment, we presented 8 practice trials chosen randomly from all 5 trial types with the exception that no trial type was played more than twice.

Data analysis. The data from Experiment 2 were processed and analyzed as in Experiment 1 except as follows. The data from the context period were analyzed in terms of the proportion of time participants heard two sounds (beeping and buzzing tones) in a similar manner as for perception of one or two streams during the test was analyzed in Experiment 1. An important difference is that the data were quantified for the entire period, even when the button was not being pressed, unlike the case for the test data in which the two alternative buttons allowed us to differentiate when participants were not responding yet versus hearing one or two streams. The average proportions of hearing two sounds during the context and test for the four different trial types with nonsilent contexts were entered into separate two-factor repeated-measures ANOVAs to test for differences depending on whether the context had a mistuned harmonic or not and whether it did or did not contain the matching 400 Hz tone. An additional one-factor repeated-measures ANOVA on the test data was performed comparing the silent condition to the four other conditions.

The data from Experiment 3 were processed and analyzed as in Experiment 2 except as follows. The average proportions of hearing two sounds during the context and test for the two different nonsilent context types were entered into separate ANOVAs. The context ANOVA tested for differences depending on whether the context had a mistuned tone or not and to which ear the critical context tone and test tones were presented. The test ANOVA included these two factors, in addition to whether the context was silent or not. This full factorial model was made possible because each trial type had its own set of silent context trials.

The data from Experiment 4 were processed and analyzed as in Experiment 2 except as follows. The average proportions of hearing two sounds during the context and test for the four different trial types with nonsilent contexts were entered into separate two-factor repeated-measures ANOVAs to test for differences depending on whether the context had a mistuned tone or not and whether it did or did not contain the matching 696 Hz tone. An additional one-factor repeated-measure ANOVA on the test data was performed comparing the silent condition to the four other conditions.

The data from Experiments 5 and 6 were processed and analyzed as in Experiment 2 except as follows. The average proportions of hearing two sounds during the context and test for the four different trial types with nonsilent contexts were entered into separate two-factor repeated-measures ANOVAs to test for differences depending on whether the context had an asynchronous tone or not and whether it did or did not contain the matching 400 Hz tone. An additional one-factor repeated-measures ANOVA on the test data was performed comparing the silent condition to the four other conditions.

Results and Discussion

Remarkably, as shown in Figures 3–7, none of the nonsilent conditions from Experiments 2–6 showed a significant facilitation of streaming compared with a silent context control condition, despite the fact that participants generally reported being very much aware of and attending to the critical context tones. This pattern of results held regardless of whether the critical context tones were made perceptually aware by mistuning the critical tones (Experiments 2–4) and presenting the critical context tones in a separate ear from the other context tones (Experiment 3), lengthening the context tones (Experiments 3 and 4), making the critical context tones asynchronous with the other context tones (Experiments 5 and 6), and using complex tones during the test sequence to make the context and test more acoustically similar (Experiments 4 and 6). The full details of the results and specific implications for all of these experiments are described below.

In Experiment 2, participants reported perceiving two sounds during the context much more in the mistuned (74.7 and 72.6% of the time for the matched and mismatched conditions, respectively) than the tuned (27.3 and 25.1% of the time for the matched and mismatched conditions, respectively) conditions, $F(1, 9) = 6.05$, $p < .05$, $\eta_p^2 = .402$. There was no significant effect of matching during the context, $F(1, 9) = 1.37$, $p = .27$, $\eta_p^2 = .132$, nor was

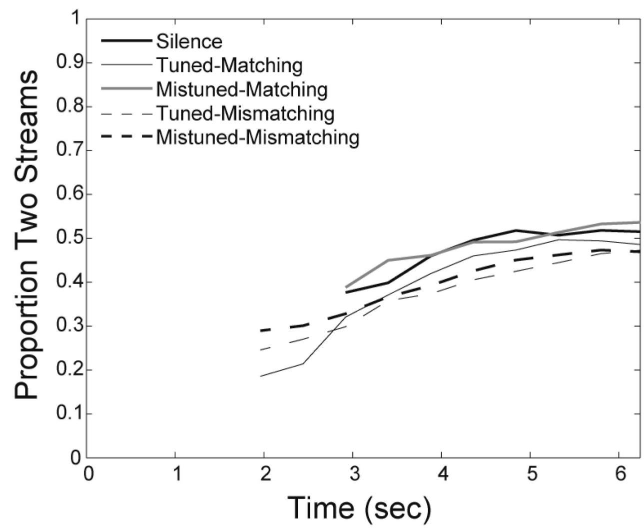

Figure 3. Time course data of streaming during the test from Experiment 2. None of the contexts facilitated streaming, even when the frequency-matching tone was mistuned from the rest of the complex tone.

there an interaction between tuning and matching, $F(1, 9) = 0.00$, $p = .982$, $\eta_p^2 = .000$. The rate of reporting two sounds in the tuned condition was higher than expected, suggesting some degree of bias to report two sounds. Figure 3 shows the time course of the "two streams" percept during the test, averaged across all participants during Experiment 2. Despite the robust sensitivity of this task to detect perception of streaming as shown in Experiment 1, the data from this experiment show a striking lack of facilitation of streaming by any of the nonsilent contexts. This was even true for the mistuned-matching condition, in which a mistuned harmonic matching the frequency of the B tone of the test, and perceived as popping out by participants, was presented during the context. The lack of streaming facilitation by the contexts was confirmed by nonsignificant main effects of tuning, $F(1, 9) = 1.33$, $p = .279$, $\eta_p^2 = .128$, and matching, $F(1, 9) = 0.98$, $p = .348$, $\eta_p^2 = .098$, and the nonsignificant interaction of these two factors, $F(1, 9) = 0.31$, $p = .594$, $\eta_p^2 = .033$. Of some note, the matching factor was marginal and in the expected direction with somewhat more streaming for matching than nonmatching conditions. However, these conditions were not above the silent condition, as confirmed by a lack of condition effect in the one-factor ANOVA, $F(4, 36) = 0.70$, $p = .594$, $\eta_p^2 = .073$. The results of this experiment show that even when a context tone is mistuned, such that it is consciously accessible and attended separately from the rest of a complex tone, it still fails to facilitate perception of streaming in a subsequent test ABA− sequence. Thus, conscious awareness of and attention to a tone is not sufficient for it to cause streaming, despite the fact that these factors may be necessary for streaming facilitation to occur. Furthermore, it is even less likely that the null results observed are attributable to peripheral interference between the critical tone and the rest of complex tone, compared with Experiment 1, because in the current experiment the critical tone was the second harmonic of a complex and the third harmonic was missing. This means that the critical tone was a full octave away from the fundamental and more than an octave away from the next highest harmonic in the

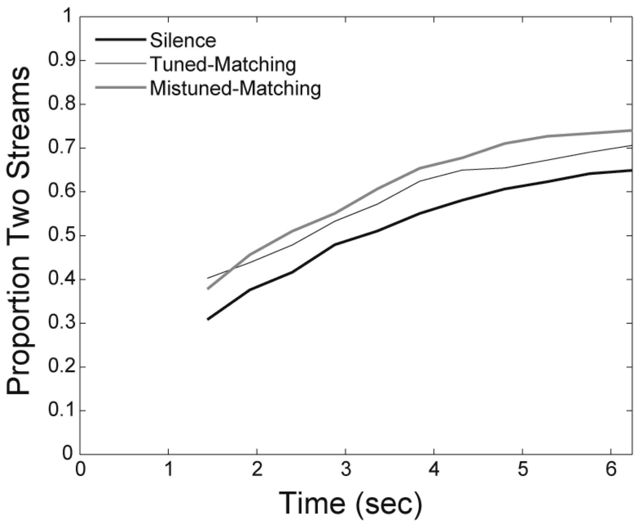

Figure 4. Time course data of streaming during the test from Experiment 3. None of the contexts facilitated streaming, even when the frequency-matching tone was mistuned from the rest of the complex tone and presented in a different ear from the rest of the complex but the same (ipsilateral) ear as the ABA− test.

mistuned-matching condition, distances that are much larger than auditory frequency-channel bandwidths (Glasberg & Moore, 1990).

In Experiment 3, participants reported perceiving two sounds during the context much more in the mistuned (93.6 and 94.5% of the time when the critical tone was in the left and right ears, respectively) than the tuned (17.4 and 25.8% of the time when the critical tone was in the left and right ears, respectively) conditions, $F(1, 11) = 69.81$, $p < .001$, $\eta_p^2 = .864$. There was a marginal effect of ear of presentation during the context, $F(1, 11) = 4.30$, $p = .06$, $\eta_p^2 = .281$, and there was no interaction between tuning and ear of presentation, $F(1, 11) = 3.10$, $p = .11$, $\eta_p^2 = .220$. Figure 4 shows the time course of the "two streams" percept, averaged across all participants during Experiment 3. Despite the presentation of the critical tone of the context in a different ear from the rest of the complex but the same ear as the ABA− test sequence, no facilitation of streaming occurred in the mistuned-matching condition. This was confirmed by the three-factor ANOVA, showing no significant main effects of tuning, $F(1, 11) = 0.08$, $p = .786$, $\eta_p^2 = .007$, or ear of presentation, $F(1, 11) = 3.07$, $p = .107$, $\eta_p^2 = .218$, and no interaction between these two factors, $F(1, 11) =$

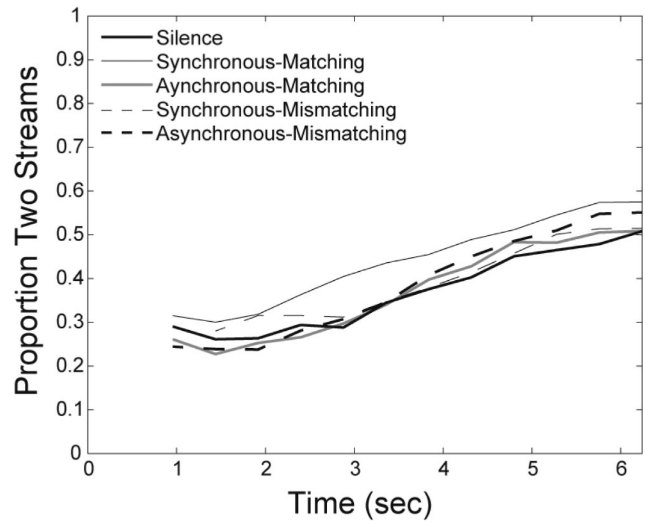

Figure 6. Time course data of streaming during the test from Experiment 5. None of the contexts facilitated streaming, even when the frequency-matching tone was asynchronous compared with the rest of the complex tone.

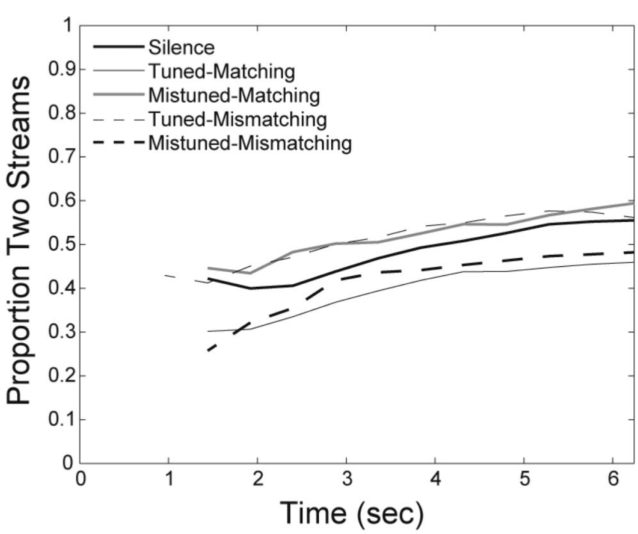

Figure 5. Time course data of streaming during the test from Experiment 4. None of the contexts facilitated streaming, even when the frequency-matching tone was mistuned from the rest of the complex tone and both the context and test patterns contained complex tones.

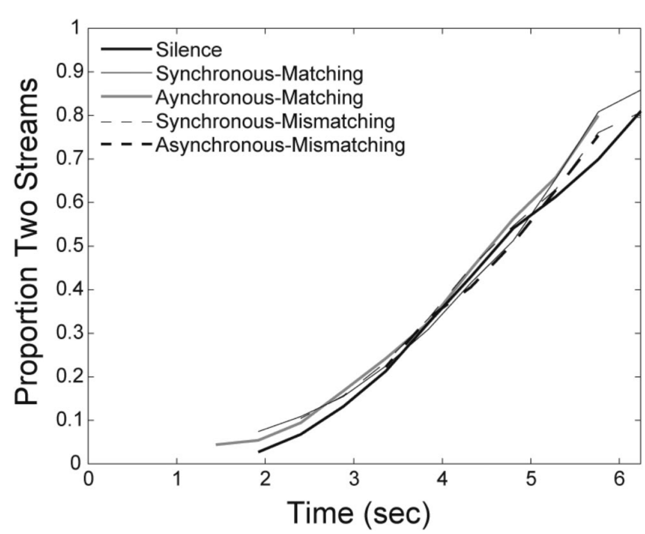

Figure 7. Time course data of streaming during the test from Experiment 6. None of the contexts facilitated streaming, even when the frequency-matching tone was asynchronous compared with the rest of the complex tone and both the context and test patterns contained complex tones.

4.04, $p = .070$, $\eta_p^2 = .269$. Furthermore, there was no main effect of whether the context was silent or not, $F(1, 11) = 3.43$, $p = .091$, $\eta_p^2 = .238$, nor did any of the factors or the interaction above interact with this effect (all ps $> .15$). This experiment thus confirmed our suspicion that the lack of facilitation of streaming when the critical context tone is consciously heard as separate from the rest of the complex context tones is not attributable to peripheral interference effects. Rather, it seems likely that a form of central interference between the critical tone and the rest of the complex prevented the critical tone from facilitating streaming in the experiments reported thus far. We will return to this issue in the General Discussion.

In Experiment 4, participants reported perceiving two sounds during the context much more in the mistuned (76.2 and 78.2% of the time for the matched and mismatched conditions, respectively) than the tuned (31.6 and 32.7% of the time for the matched and mismatched conditions, respectively) conditions, $F(1, 17) = 26.23$, $p < .001$, $\eta_p^2 = .607$. There was no significant effect of matching during the context, $F(1, 17) = 1.17$, $p = .29$, $\eta_p^2 = .064$, nor was there an interaction between tuning and matching, $F(1, 17) = 0.04$, $p = .851$, $\eta_p^2 = .002$. Figure 5 shows the time course of the "two streams" percept, averaged across all participants during Experiment 4. As in the previous experiments, none of the complex-tone contexts facilitated perception of streaming, not even the mistuned-matching context. This null finding was supported by nonsignificant effects of tuning, $F(1, 17) = 0.49$, $p = .494$, $\eta_p^2 = .028$, and matching, $F(1, 17) = 0.17$, $p = .683$, $\eta_p^2 = .010$. Although there was a significant interaction between these factors, $F(1, 17) = 7.73$, $p < .025$, $\eta_p^2 = .313$, the pattern of differences were not consistent with the frequency-dependent facilitation of streaming observed with simple tones in Experiment 1. In particular, although the mistuned-matching condition resulted in slightly more streaming compared with the silent context, so did the tuned-mismatching condition. The ANOVA including the silent context showed no main effect of condition, $F(4, 68) = 1.99$, $p = .134$, $\eta_p^2 = .105$. The results of this experiment thus rule out that the lack of facilitation of streaming from a matching tone embedded as a tuned or mistuned harmonic in a complex tone is attributable to the change from a complex-tone sequence during the context to a simple-tone sequence during the test. Rather, it seems more likely that some form of interference is occurring, such that a simple tone accompanied by a concurrent complex tone cannot facilitate streaming of a later ABA− sequence.

In Experiment 5, participants reported perceiving two sounds during the context much more in the asynchronous (68.4 and 69.9% of the time for the matched and mismatched conditions, respectively) than the synchronous (14.2 and 20.4% of the time for the matched and mismatched conditions, respectively) conditions, $F(1, 10) = 30.56$, $p < .001$, $\eta_p^2 = .753$. There was no significant effect of matching during the context, $F(1, 10) = 1.68$, $p = .224$, $\eta_p^2 = .144$, nor was there an interaction between synchrony and matching, $F(1, 10) = 0.64$, $p = .442$, $\eta_p^2 = .060$. Figure 6 shows the time course of the "two streams" percept, averaged across all participants during Experiment 5. As in the previous experiments, none of the complex-tone contexts facilitated perception of streaming, not even the asynchronous-matching context. This null finding was supported by nonsignificant effects of synchrony, $F(1, 10) = 1.44$, $p = .257$, $\eta_p^2 = .126$, and matching, $F(1, 10) = 1.80$, $p = .209$, $\eta_p^2 = .153$, and a nonsignificant interaction between these factors, $F(1, 10) = 2.24$, $p = .166$, $\eta_p^2 = .183$. The ANOVA including the silent context showed no main effect of condition, $F(4, 40) = 1.57$, $p = .202$, $\eta_p^2 = .135$. The results of this experiment thus rule out that the lack of facilitation of streaming facilitation from a matching tone is particular to using mistuning to form a separate auditory object from the critical tone. Here, we have shown that when the critical tone is heard as a separate auditory object as a result of having an asynchronous onset it still does not promote perception of streaming in the test ABA− pattern.

In Experiment 6, participants reported perceiving two sounds during the context much more in the asynchronous (74.6 and 70.3% of the time for the matched and mismatched conditions, respectively) than the synchronous (27.3 and 22.0% of the time for the matched and mismatched conditions, respectively) conditions, $F(1, 21) = 62.75$, $p < .001$, $\eta_p^2 = .749$. There was a significant effect of matching during the context, $F(1, 21) = 4.96$, $p < .05$, $\eta_p^2 = .191$, attributable to slightly more perception of two objects in the matching conditions, but there was no interaction between synchrony and matching, $F(1, 21) = 0.09$, $p = .774$, $\eta_p^2 = .004$. Figure 7 shows the time course of the "two streams" percept, averaged across all participants during Experiment 6. As in the previous experiments, none of the complex-tone contexts facilitated perception of streaming, not even the asynchronous-matching context. This null finding was supported by nonsignificant effects of synchrony, $F(1, 21) = 0.05$, $p = .831$, $\eta_p^2 = .002$, and matching, $F(1, 21) = 0.27$, $p = .612$, $\eta_p^2 = .012$, and a nonsignificant interaction between these factors, $F(1, 21) = 0.04$, $p = .847$, $\eta_p^2 = .002$. The ANOVA including the silent context showed no main effect of condition, $F(4, 84) = 0.31$, $p = .825$, $\eta_p^2 = .015$. The results of this experiment thus rule out that the lack of facilitation of streaming from an asynchronous matching context tone is attributable to the change from hearing complex tones during the context to simple tones during the test.

Experiment 7

The preceding experiments' findings suggest that the facilitating effect of a matching context tone on streaming, though large when using single-tone contexts, is a rather fragile effect that can be prevented by a number of situations in which other tones are present. The purpose of this experiment was to better delineate the boundary conditions in which streaming facilitation occurs. We were particularly interested to see whether nontone sounds would interfere with the critical context tone's facilitating influence on streaming during the test. We therefore presented bandpass noises and speech samples that were synchronous with the critical matching tone, as well as continuous speech that had onsets and offsets that were uncorrelated with the tone onsets and offsets during the context. As in Experiment 1, we also included single-tone contexts in addition to a silent-context condition, for comparison purposes.

Method

Participants. Ten adults with normal self-reported hearing (four men and six women, age range = 20–34 years, mean age = 26.10 years) from the University of Nevada, Las Vegas subject pool or Psychology Graduate Program participated after giving written informed consent according to the guidelines of the University's Office for the Protection of Research Subjects.

Stimuli and procedure. The stimuli and procedure were the same as in Experiment 2, except as follows. As shown in Figure 1, the test stimuli were presented in an ABAB pattern (i.e., in an isochronous fashion) rather than an ABA− pattern. The test tones were 100 ms in duration, including 10-ms rise/fall times with linear ramps. The frequency of the A tones was fixed at 282 Hz and the frequency of the B tones was fixed at 400 Hz. This corresponds to an A-B Δf of 6 semitones.

All context sequences contained a 400 Hz simple tone that matched the frequency of the B test tone. The tones were repeated with a constant 120-ms SOA, were gated on and off with 10-ms linear ramps, and in all except one condition had a duration of 100 ms. Context tones were presented concurrently with one of five possible background sounds (available for listening at http://faculty.unlv.edu/jsnyder/example_stimuli.zip). 1) The *simple* condition contained no background, similar to the *simple-matching* condition in Experiment 1 except that the critical context tone matched the frequency of the B test tone rather than the A test tone. 2) The *asynchronous* condition contained a background of a repeating complex tone with a fundamental frequency of 200 Hz and integer-multiple harmonics of 600, 800, and 1000 Hz. The complex tone was 100 ms in duration, including 10-ms rise/fall times with linear ramps and a constant 120-ms SOA. The critical context tone was embedded in the complex tone as the second harmonic. Importantly, the context tones had an onset that was 20 ms later and offset that was 20 ms earlier than the other tones. As a result, the context tone was 60 ms rather than 100 ms as in the other conditions. This condition is similar to the *asynchronous-matching* condition of Experiments 5 and 6 expect that the critical tone had an asynchronous onset *and* offset. 3) The *synchronous speech* condition contained a background sound, consisting of a native-English speaking male speaker repeating the syllable /ba/ recorded in a sound attenuated booth. The beginning and end portion of the recorded sample was discarded so that the remaining syllable was 100 ms with 10-ms rise/fall times with linear ramps. As a result, the remaining stimulus sounded more like /a/ than it did /ba/. This stimulus was presented synchronously with the context tones. 4) The *continuous speech* condition contained a background sound consisting of the same male speaker as in the synchronous speech condition reading the sentence "The discrimination of electrical stimuli by a deaf person is generally much less acute than the discrimination of acoustical stimuli by a normally hearing person" from a textbook (Moore, 2003). The beginning and end portion of the recorded sample was discarded so that the remaining sentence was 6.72 s. The remaining sentence was ". . .lectrical stimuli by a deaf person is generally much less acute than the discrimination of acoustical stimuli by a nor. . ." The onset and offset of the remaining sentence were modified to have 10-ms rise/fall times with linear ramps. Of the remaining word onsets and offsets, only three word onsets coincided with the onset of a context tone and only five word offsets coincided with the offset of a context tone within a 10-ms time window. As shown in Figure 8, the *synchronous speech* and *continuous speech* differed in many respects. Therefore, differences in streaming facilitation between these two conditions may be ascribed to any one or combination of these differences. Notable differences between the two conditions include a high correlation between speech onsets and critical tone onsets in the *synchronous speech* condition, the increased temporal variation (e.g., formant transitions and variable breaks between speech utterances) in the *continuous speech* condition, and the lack of intelligibility of the *synchronous speech* condition. 5) The *bandpass noise* condition used background sounds consisting of noise generated in MATLAB. The bandpass noise was 100 ms with 10-ms rise/fall times with linear ramps. The noise was played synchronously with the context tones. The *synchronous speech* and *continuous speech* were filtered using a rectangular band such that frequencies >200 Hz and <600 Hz were removed. Furthermore, frequencies >1000 Hz were additionally removed in the *bandpass noise* to match the spectral range of the *asynchronous-matching* condition. Frequencies <200 Hz were not filtered out. 6) The *silence* condition was included in which the context period was filled with 6.72 s of silence, which served as a baseline condition. Figure 8 shows spectrograms of the complex context stimuli that were played with the critical tones.

Four blocks of 30 trials were presented (20 of each of the six trial types). The different types of trials were randomly intermingled within a block. Before the experiment, we presented 6 prac-

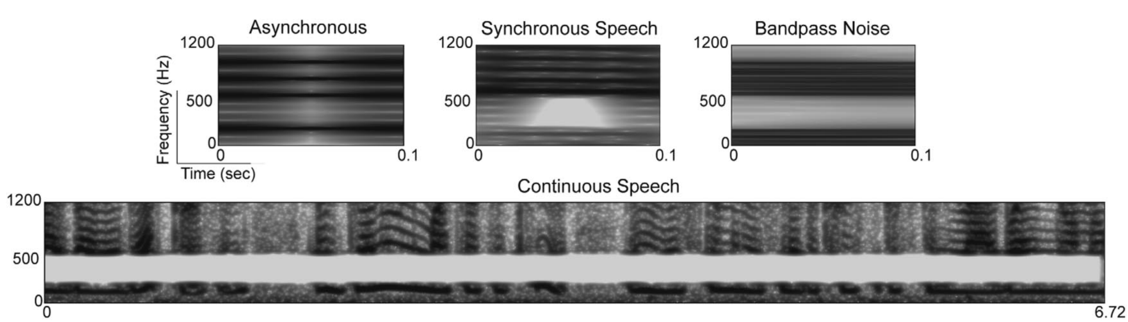

Figure 8. Spectrograms for complex context stimuli in Experiment 7. Darker portions of the spectrogram indicate higher amounts of energy. The asynchronous, synchronous speech, and bandpass noise stimuli were presented for 100 ms in precise temporal relationships to the critical tone. The continuous speech stimulus was simply played throughout the context and the critical tones were played at times independently of events in the speech stimulus. Note that the spectrograms of the asynchronous, synchronous speech, and bandpass noise stimuli have edge artifacts because of their very brief nature. Their actual spectral features have much sharper boundaries in frequency and time, as verified with longer versions of these stimuli.

tice trials chosen randomly from all 6 trial types with the exception that no trial type was played more than once.

Data analysis. The data were processed and analyzed as in Experiment 2 except as follows. The average proportions of hearing two sounds during the context and test for all six conditions were entered into separate one-factor repeated-measure ANOVAs to test for differences between the conditions. A set of simple contrasts was also performed comparing proportion of hearing two sounds during the test for all conditions with a nonsilent context to the condition with a silent context.

Results and Discussion

Participants reported hearing the critical context tone for most of the context (>80% of the time) when the tone was present but not when it was absent (i.e., except for the silence condition), resulting in a main effect of condition, $F(5, 45) = 322.28$, $p < .001$, $\eta_p^2 = .973$. However, there was no main effect of condition when excluding the silence condition, $F(4, 36) = 0.82$, $p = .439$, $\eta_p^2 = .084$, indicating that the nonsilent conditions did not differ from each other in terms of audibility of the critical context tone.

Figure 9 shows the time course of the "two streams" percept during the test, averaged across all participants during Experiment 7. As shown in the previous experiments, a context with a matching pure tone and no other accompanying tones resulted in a large amount of streaming facilitation, but contexts with complex tones did not facilitate perception of streaming (compared with the silent context), $F(1, 9) = .40$, $p = .545$, $\eta_p^2 = .042$, even when there were onset and offset asynchronies between the critical tone and the rest of the complex tone. Furthermore, contexts with synchronous speech did not facilitate perception of streaming (compared with the silent context), $F(1, 9) = 2.71$, $p = .134$, $\eta_p^2 = .232$. Unlike the previous experiments, however, the continuous speech context and the bandpass noise context showed a large amount of streaming facilitation, although less than the context with the pure tone by itself. This pattern of findings resulted in a main effect of condition, $F(5, 45) = 7.59$, $p < .001$, $\eta_p^2 = .457$, and significantly more streaming in the pure, continuous speech, and bandpass noise conditions compared with the silence condition, $F(1, 9) = 18.35$, 11.02, and 6.00, $p < .005$, .01, and .05, $\eta_p^2 = .671$, .550, and .400. This experiment shows that although facilitation of streaming is quite easy to disrupt when other tones are presented at the same time as the critical context tones, facilitation might be more likely to occur when the accompanying sounds are distinct in terms of timbre (as with the noise and continuous speech contexts) and/or have onsets and offsets that are not correlated with the critical tones (as with the continuous speech context).

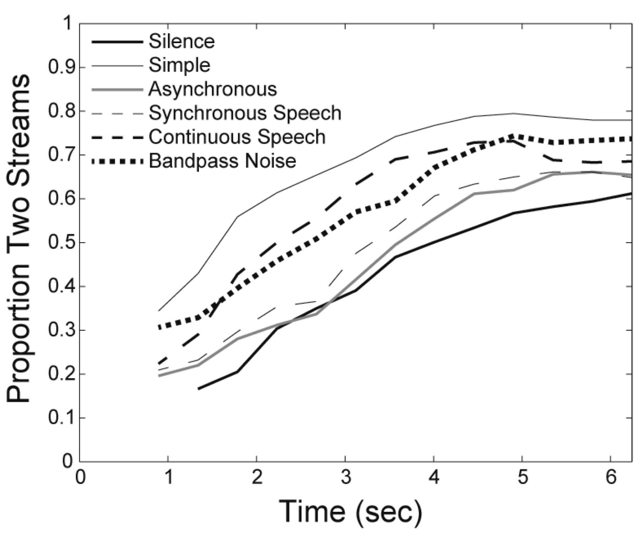

Figure 9. Time course data of streaming during the test from Experiment 7. None of the contexts facilitated streaming, except when using continuous speech or bandpass noise.

General Discussion

The experiments described here showed how extra sounds can interfere with the facilitation of streaming that normally occurs as a result of context tones that match the frequency of one of the tones in a following test sequence of alternating A and B tones. In Experiment 1, we found that when the context tones are made inaccessible to attention and awareness as a result of being harmonic partials in complex tones, they do not cause facilitation of streaming during the test. We also found that even when the context tones are mistuned (Experiments 2–4) or asynchronous (Experiments 5–7) relative to the rest of the complex tone, they still do not facilitate streaming during the test. This is true even when the context tone is presented to the opposite ear relative to the rest of the complex tone (Experiment 3) and even when timbre changes from context to test are ruled out as interfering with facilitation (Experiments 4 and 6). The only situations in which an accompanying sound did not prevent facilitation of streaming were when the accompanying sound was continuous speech or bandpass noise (Experiment 7). Below, we discuss how these results update our understanding of auditory stream segregation.

Implications for Theories of Streaming

Our findings are consistent with the views put forward by several reviews that a comprehensive account of stream segregation must incorporate both low-level and high-level processes (Alain, 2007; Bregman, 1990; Moore & Gockel, 2002; Snyder & Alain, 2007; Snyder et al., 2012). In particular, the finding that a context tone that matches the frequency of one of the test tones is not sufficient to facilitate streaming when it is not accessible to attention and awareness argues for an important role of high-level processes in streaming. We assume that the context tones that were embedded in complex tones nevertheless evoked substantial frequency-specific adaptation in peripheral auditory neurons tuned to its frequency, because of the relatively wide frequency separation from other components of the complex tones (Glasberg & Moore, 1990). Importantly, such frequency-specific adaptation in the brainstem and auditory cortex has been proposed to be a primary mechanism underlying stream segregation (Micheyl et al., 2005; Pressnitzer et al., 2008). Thus, although such low-level neural adaptation may be necessary for streaming to occur, it might not be sufficient, contrary to theories that have made strong claims about the exclusively peripheral nature of bottom-up stream segregation processes (Anstis & Saida, 1985; Beauvois & Meddis, 1996; Hartmann & Johnson, 1991; Van Noorden, 1975).

Instead, alternative theories of stream segregation that require the higher-level representation of different possible perceptual organizations might be required to fully account for streaming (Bregman, 1990; Denham & Winkler, 2006). For example, Bregman (1990) explains streaming as a process in which the auditory system initially assumes that there is one sound source and only after accumulating enough evidence does it consider the possibility that there might be two sound sources. This theory can explain the facilitation of streaming by a repeating-frequency context tone as the priming of the possibility that there are two sound sources. Theories that posit the existence of competing representations are also consistent with the recent observation of multistable switching between one- and two-stream percepts during streaming after the initial segregation process is complete (Kondo & Kashino, 2009; Pressnitzer & Hupé, 2006), similar to the switching that occurs with other auditory paradigms such as verbal transformation (Ditzinger, Tuller, & Kelso, 1997; Pitt & Shoaf, 2002; Warren, 1968), ambiguous visual figures (Leopold & Logothetis, 1999; Long & Toppino, 2004), and ambiguous tactile patterns (Carter, Konkle, Wang, Hayward, & Moore, 2008).

Another possible explanation of our results is that some form of interference between the matching context tone and the other simple tones prevented the matching tone from facilitating streaming. One potential mechanism for such interference that can occur at multiple levels of the central auditory system is wide-band inhibition, whereby responses to a particular target tone can be suppressed by spectrally distant tones, even when the interacting tones are presented to different ears (Bartlett & Wang, 2005; Bleeck, Ingham, Verhey, & Winter, 2008; Shore, Sumner, Bledsoe, & Lu, 2003; for an application to auditory scene analysis see Roberts & Holmes, 2007). Informational masking is another way of conceptualizing interference that occurs at central levels of processing and can involve interactions between sounds of very distant frequencies (Kidd, Mason, & Richards, 2003). A recent study showed that informational unmasking, as measured by the ability to detect an isochronous series of tones amid a background of tones of distant frequencies and random timing, occurred concomitantly with neuromagnetic activation at the level of secondary auditory cortex, but not in primary auditory cortex (Gutschalk, Micheyl, & Oxenham, 2008). However, it is difficult to conclude that the interference observed in the current study between tones of different frequency was attributable to informational masking because even when the critical tone was unmasked by various means, it still failed to facilitate streaming in the presence of other tones.

Fragility of Streaming

To our knowledge, all previous experiments that have demonstrated streaming facilitation have presented a series of isolated context tones (e.g., Beauvois & Meddis, 1997; Haywood & Roberts, in press; Rogers & Bregman, 1993, 1998) or a series of alternating A and B tones (e.g., Anstis & Saida, 1985; Bregman, 1978; Carlyon et al., 2001; Snyder et al., 2008), before an alternating pattern of A and B test tones. Although these previous studies all show large facilitative effects of a frequency-matching context on streaming, Experiments 2–7 of the current study suggest that streaming facilitation can be quite a fragile process even when the critical context tones are clearly heard out as separate auditory objects by virtue of being mistuned, asynchronous, distinct in timbre, or presented in the other ear relative to the rest of the context. In particular, the presence of other simultaneous context tones strongly suppressed the facilitation of streaming, which we thought might occur when the frequency-matching context tones were mistuned, asynchronous, distinct in timbre, or presented in a different ear. An important point to make, however, is that our findings may not generalize to streaming facilitation when alternating A and B tones are used in the context, as opposed to the single-frequency critical tones used in the current study (cf. Thompson et al., 2011) because studies have not shown that streaming facilitation with these two types of contexts relies on similar mechanisms. Furthermore, there is recent evidence that the two types of facilitation might rely on different mechanisms, based on the observation of different temporal dynamics of the two effects (Haywood & Roberts, in press).

Although it is surprising that a perceptually segregated simple tone does not facilitate streaming for a later ABA− pattern, our findings are consistent with a number of other studies demonstrating the fragility of streaming facilitation. For example, a number of studies using various behavioral and neurophysiological methods have suggested that streaming is strongly diminished when attention is diverted away from ABA− tones (Carlyon et al., 2001; Elhilali, Xiang, et al., 2009; Snyder et al., 2006; Thompson et al., 2011; but see Macken, Tremblay, Houghton, Nicholls, & Jones, 2003; Sussman et al., 2007). Similarly, even briefly switching attention or the presence of brief silent gaps strongly resets streaming (Cusack et al., 2004). Studies have also shown that changes in the frequency, intensity, or perceived location, and self-induced motion, can reset streaming (Anstis & Saida, 1985; Roberts et al., 2008; Rogers & Bregman, 1998; Kondo et al., 2012). Thus, it is possible that acoustic changes from the context to the test may have reset streaming, although our experiments using complex tones during the context and the test (Experiment 4 and 6) do not support this conclusion because streaming facilitation was still disrupted to a similar extent despite the increased acoustic similarity of the context and test stimuli compared with our experiments that used simple-tone test stimuli.

Implications for Real-World Listening

An important question that arises from the idea of the fragility of streaming facilitation is what should be expected in real-world listening situations. For example, if a series of notes played by a musical instrument or words spoken by a person is accompanied by other simultaneous sounds, can the series of notes or words facilitate the segregation of the instrument or speaker when later accompanied by different sounds than were present earlier? Some hints to a possible answer arise from a recent study by McDermott and colleagues (2011) using a very different type of segregation paradigm. In this study, a spectro-temporally dynamic noise target pattern was repeated a number of times in a trial, each time mixed with other spectro-temporal noises. As long as each repetition of the target pattern was mixed with a different distracter pattern, the repetition facilitated perceptual segregation of the target from the rest of the mixture, as measured by listeners' ability to recognize the identity of the target stimulus when presented alone following the series of mixtures. This is also consistent with the results of the continuous speech context condition of Experiment 7, because the speech onsets and offsets were not temporally correlated with the critical pure tones. In this condition, almost as much streaming facilitation occurred as when the context was the critical pure tone by itself. It is also possible that presenting a different complex tone along with each repetition of the frequency-matching context tone might enable the context tones to facilitate streaming in the subsequent ABA− test. Such a situation might better model real-world listening situations because it is unlikely that different sound sources will produce similar sounds in a temporally correlated fashion.

A final consideration related to real-world listening is that the sounds we used in the current study are acoustically and semantically unnatural. Perhaps in real-world situations, such as when listening to music, speech, and environmental sounds that have rich spectral and temporal structure and various degrees of meaning or familiarity, streaming facilitation is much more robust to the factors we discussed above that make it appear so fragile in laboratory experiments using static simple tones. Alternatively, it is possible that streaming facilitation is not as relevant in real-world situations because of the less ambiguous nature of segregation when using familiar sounds. Surprisingly, despite the several decades of research on streaming using simple ABA− patterns, we know very little about how the phenomena we have identified in the laboratory apply to real-world sounds and real-world situations. Thus, it is important to perform more experiments using complex and meaningful sounds examining the various factors that facilitate and inhibit streaming, as has been done somewhat sporadically over the years (e.g., Billig, Davis, Deeks, Monstrey, & Carlyon, 2013; Dorman, Cutting, & Raphael, 1975; Dowling, Lung, & Herrbold, 1987; Gaudrain, Grimault, Healy, & Bera, 2008; Hartmann & Johnson, 1991).

References

Alain, C. (2007). Breaking the wave: Effects of attention and learning on concurrent sound perception. *Hearing Research, 229,* 225–236. doi:10.1016/j.heares.2007.01.011

Alain, C., Schuler, B. M., & McDonald, K. L. (2002). Neural activity associated with distinguishing concurrent auditory objects. *Journal of the Acoustical Society of America, 111,* 990–995. doi:10.1121/1.1434942

Anstis, S., & Saida, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 257–271. doi:10.1037/0096-1523.11.3.257

Bartlett, E. L., & Wang, X. (2005). Long-lasting modulation by stimulus context in primate auditory cortex. *Journal of Neurophysiology, 94,* 83–104. doi:10.1152/jn.01124.2004

Beauvois, M. W., & Meddis, R. (1996). Computer simulation of auditory stream segregation in alternating-tone sequences. *Journal of the Acoustical Society of America, 99,* 2270–2280. doi:10.1121/1.415414

Beauvois, M. W., & Meddis, R. (1997). Time decay of auditory stream biasing. *Perception & Psychophysics, 59,* 81–86. doi:10.3758/BF03206850

Billig, A. J., Davis, M. H., Deeks, J. M., Monstrey, J., & Carlyon, R. P. (2013). Lexical influences on auditory streaming. *Current Biology, 23,* 1585–1589. doi:10.1016/j.cub.2013.06.042

Bleeck, S., Ingham, N. J., Verhey, J. L., & Winter, I. M. (2008). Rebound depolarization in single units of the ventral cochlear nucleus: A contribution to grouping by common onset? *Neuroscience, 154,* 139–146. doi:10.1016/j.neuroscience.2008.03.020

Bregman, A. S. (1978). Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 380–387. doi:10.1037/0096-1523.4.3.380

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

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. doi:10.1037/h0031163

Bullier, J. (2001). Feedback connections and conscious vision. *Trends in Cognitive Sciences, 5,* 369–370.

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 115–127. doi:10.1037/0096-1523.27.1.115

Carlyon, R. P., Plack, C. J., Fantini, D. A., & Cusack, R. (2003). Cross-modal and non-sensory influences on auditory streaming. *Perception, 32,* 1393–1402. doi:10.1068/pp.5035

Carter, O., Konkle, T., Wang, Q., Hayward, V., & Moore, C. (2008). Tactile rivalry demonstrated with an ambiguous apparent-motion quartet. *Current Biology, 18,* 1050–1054. doi:10.1016/j.cub.2008.06.027

Cusack, R., Deeks, J., Aikman, G., & Carlyon, R. P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 643–656. doi:10.1037/0096-1523.30.4.643

Cusack, R., & Roberts, B. (2000). Effects of differences in timbre on sequential grouping. *Perception & Psychophysics, 62,* 1112–1120. doi:10.3758/BF03212092

Denham, S. L., & Winkler, I. (2006). The role of predictive models in the formation of auditory streams. *Journal of Physiology (Paris), 100,* 154–170. doi:10.1016/j.jphysparis.2006.09.012

Ditzinger, T., Tuller, B., & Kelso, J. A. S. (1997). Temporal patterning in an auditory illusion: The verbal transformation effect. *Biological Cybernetics, 77,* 23–30. doi:10.1007/s004220050363

Dorman, M. F., Cutting, J. E., & Raphael, L. J. (1975). Perception of temporal order in vowel sequences with and without formant transitions. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 121–129. doi:10.1037/0096-1523.1.2.121

Dowling, W. J., Lung, K. M. T., & Herrbold, S. (1987). Aiming attention in pitch and time in the perception of interleaved melodies. *Perception & Psychophysics, 41,* 642–656. doi:10.3758/BF03210496

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. doi:10.1016/j.neuron.2008.12.005

Elhilali, M., Xiang, J., Shamma, S. A., & Simon, J. Z. (2009). Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. *PLoS Biology, 7,* e1000129. doi:10.1371/journal.pbio.1000129

Fowler, C. A., & Rosenblum, L. D. (1990). Duplex perception: A comparison of monosyllables and slamming doors. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 742–754. doi:10.1037/0096-1523.16.4.742

Gaudrain, E., Grimault, N., Healy, E. W., & Bera, J. C. (2008). Streaming of vowel sequences based on fundamental frequency in a cochlear-implant simulation. *Journal of the Acoustical Society of America, 124,* 3076–3087. doi:10.1121/1.2988289

Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research, 47,* 103–138. doi:10.1016/0378-5955(90)90170-T

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. doi:10.1121/1.1452740

Gutschalk, A., Micheyl, C., & Oxenham, A. J. (2008). Neural correlates of auditory perceptual awareness under informational masking. *PLoS Biology, 6,* e138. doi:10.1371/journal.pbio.0060138

Hartmann, W. M., & Johnson, D. (1991). Stream segregation and peripheral channeling. *Music Perception, 9,* 155–183. doi:10.2307/40285527

Haywood, N. R., & Roberts, B. (in press). Build-up of auditory stream segregation induced by tone sequences of constant or alternating frequency and the resetting effects of single deviants. *Journal of Experimental Psychology: Human Perception and Performance*. doi:10.1037/a0032562

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron, 36,* 791–804. doi:10.1016/S0896-6273(02)01091-7

Kidd, G., Jr., Mason, C. R., & Richards, V. M. (2003). Multiple bursts, multiple looks, and stream coherence in the release from informational masking. *Journal of the Acoustical Society of America, 114,* 2835–2845. doi:10.1121/1.1621864

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends in Cognitive Sciences, 11,* 16–22. doi:10.1016/j.tics.2006.10.012

Kondo, H. M., & Kashino, M. (2009). Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. *The Journal of Neuroscience, 29,* 12695–12701. doi:10.1523/JNEUROSCI.1549-09.2009

Kondo, H. M., Pressnitzer, D., Toshima, I., & Kashino, M. (2012). Effects of self-motion on auditory scene analysis. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 109,* 6775–6780. doi:10.1073/pnas.1112852109

Lamme, V. A. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences, 7,* 12–18. doi:10.1016/S1364-6613(02)00013-X

Leopold, D. A., & Logothetis, N. K. (1999). Multistable phenomena: Changing views in perception. *Trends in Cognitive Sciences, 3,* 254–264. doi:10.1016/S1364-6613(99)01332-7

Long, G. M., & Toppino, T. C. (2004). Enduring interest in perceptual ambiguity: Alternating views of reversible figures. *Psychological Bulletin, 130,* 748–768. doi:10.1037/0033-2909.130.5.748

Macken, W. J., Tremblay, S., Houghton, R. J., Nicholls, A. P., & Jones, D. M. (2003). Does auditory streaming require attention? Evidence from attentional selectivity in short-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 43–51. doi:10.1037/0096-1523.29.1.43

McDermott, J. H., Wrobleski, D., & Oxenham, A. J. (2011). Recovering sound sources from embedded repetition. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 108,* 1188–1193. doi:10.1073/pnas.1004765108

Micheyl, C., Tian, B., Carlyon, R. P., & Rauschecker, J. P. (2005). Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron, 48,* 139–148. doi:10.1016/j.neuron.2005.08.039

Moore, B. C. J. (2003). *An introduction to the psychology of hearing*. San Diego, CA: Academic Press.

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

Moore, B. C. J., Peters, R. W., & Glasberg, B. R. (1985). Thresholds for the detection of inharmonicity in complex tones. *Journal of the Acoustical Society of America, 77,* 1861–1867. doi:10.1121/1.391937

Pitt, M. A., & Shoaf, L. (2002). Linking verbal transformations to their causes. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 150–162. doi:10.1037/0096-1523.28.1.150

Pressnitzer, D., & Hupé, J. M. (2006). Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Current Biology, 16,* 1351–1357. doi:10.1016/j.cub.2006.05.054

Pressnitzer, D., Sayles, M., Micheyl, C., & Winter, I. M. (2008). Perceptual organization of sound begins in the auditory periphery. *Current Biology, 18,* 1124–1128. doi:10.1016/j.cub.2008.06.053

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. doi:10.1037/0096-1523.34.4.992

Roberts, B., & Holmes, S. D. (2007). Contralateral influences of wideband inhibition on the effect of onset asynchrony as a cue for auditory grouping. *Journal of the Acoustical Society of America, 121,* 3655–3665. doi:10.1121/1.2721874

Rogers, W. L., & Bregman, A. S. (1993). An experimental evaluation of three theories of auditory stream segregation. *Perception & Psychophysics, 53,* 179–189. doi:10.3758/BF03211728

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. doi:10.3758/BF03206171

Shinn-Cunningham, B. G., Lee, A. K., & Oxenham, A. J. (2007). A sound element gets lost in perceptual competition. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 104,* 12223–12227. doi:10.1073/pnas.0704641104

Shore, S. E., Sumner, C. J., Bledsoe, S. C., & Lu, J. (2003). Effects of contralateral sound stimulation on unit activity of ventral cochlear nucleus neurons. *Experimental Brain Research, 153,* 427–435. doi:10.1007/s00221-003-1610-6

Snyder, J. S., & Alain, C. (2007). Toward a neurophysiological theory of auditory stream segregation. *Psychological Bulletin, 133,* 780–799. doi:10.1037/0033-2909.133.5.780

Snyder, J. S., Alain, C., & Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *Journal of Cognitive Neuroscience, 18,* 1–13. doi:10.1162/089892906775250021

Snyder, J. S., Carter, O. L., Hannon, E. E., & Alain, C. (2009). Adaptation reveals multiple levels of representation in auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 1232–1244. doi:10.1037/a0012741

Snyder, J. S., Carter, O. L., Lee, S.-K., Hannon, E. E., & Alain, C. (2008). Effects of context on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1007–1016. doi:10.1037/0096-1523.34.4.1007

Snyder, J. S., Gregg, M. K., Weintraub, D. M., & Alain, C. (2012). Attention, awareness, and the perception of auditory scenes. *Frontiers in Psychology, 3,* 15. doi:10.3389/fpsyg.2012.00015

Snyder, J. S., Holder, W. T., Weintraub, D. M., Carter, O. L., & Alain, C. (2009). Effects of prior stimulus and prior perception on neural correlates of auditory stream segregation. *Psychophysiology, 46,* 1208–1215. doi:10.1111/j.1469-8986.2009.00870.x

Snyder, J. S., & Weintraub, D. M. (2011). Pattern specificity in the effect of prior Δf on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 1649–1656. doi:10.1037/a0023098

Snyder, J. S., & Weintraub, D. M. (2013). Loss and persistence of implicit memory for sound: Evidence from auditory stream segregation context effects. *Attention, Perception, & Psychophysics, 75,* 1059–1074. doi:10.3758/s13414-013-0460-y

Sussman, E. S., Horvath, J., Winkler, I., & Orr, M. (2007). The role of attention in the formation of auditory streams. *Perception & Psychophysics, 69,* 136–152. doi:10.3758/BF03194460

Thompson, S. K., Carlyon, R. P., & Cusack, R. (2011). An objective measurement of the build-up of auditory streaming and of its modulation by attention. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 1253–1262. doi:10.1037/a0021925

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136. doi:10.1016/0010-0285(80)90005-5

Van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Eindhoven University of Technology, Eindhoven.

Vliegen, J., Moore, B. C. J., & 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. doi:10.1121/1.427140

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. doi:10.1121/1.424503

Warren, R. M. (1968). Verbal transformation effect and auditory perceptual mechanisms. *Psychological Bulletin, 70,* 261–270. doi:10.1037/h0026275

Received December 13, 2012
Revision received September 18, 2013
Accepted September 18, 2013 ■