Journal of Experimental Psychology:
Human Perception and Performance
2008, Vol. 34, No. 4, 1007–1016

0096-1523/08/$12.00 DOI: 10.1037/0096-1523.34.4.1007

Effects of Context on Auditory Stream Segregation

Joel S. Snyder
Veterans Affairs Boston Healthcare System/Harvard Medical School and Harvard University

Olivia L. Carter
Harvard University

Suh-Kyung Lee
Massachusetts Institute of Technology

Erin E. Hannon
Harvard University

Claude Alain
Baycrest Centre for Geriatric Care

The authors examined the effect of preceding context on auditory stream segregation. Low tones (A), high tones (B), and silences (−) were presented in an ABA− pattern. Participants indicated whether they perceived 1 or 2 streams of tones. The A tone frequency was fixed, and the B tone was the same as the A tone or had 1 of 3 higher frequencies. Perception of 2 streams in the current trial increased with greater frequency separation between the A and B tones (Δf). Larger Δf in previous trials modified this pattern, causing less streaming in the current trial. This occurred even when listeners were asked to bias their perception toward hearing 1 stream or 2 streams. The effect of previous Δf was not due to response bias because simply perceiving 2 streams in the previous trial did not cause less streaming in the current trial. Finally, the effect of previous Δf was diminished, though still present, when the silent duration between trials was increased to 5.76 s. The time course of this context effect on streaming implicates the involvement of auditory sensory memory or neural adaptation.

Keywords: auditory scene analysis, auditory sensory memory, neural adaptation

Real-world behaviors occur in a rich context in which recent experience can have a large influence on subsequent perception, cognition, and action. Effects of context can arise from a number of different types of mechanisms, such as sensory or perceptual adaptation, response bias, attention, learning, and intrinsic dynamics (e.g., 1/f processes; Gilden, 2001). Context effects demonstrate that at many levels, processes in the nervous system are highly dependent on previous history (for reviews, see Fecteau & Munoz, 2003; Grill-Spector, Henson, & Martin, 2006). Auditory processing may be particularly sensitive to context because stimuli, such as speech and music, often require integration of information over relatively long periods of time (e.g., up to several seconds for sentences and melodies). Examples of the generality of context effects in auditory processing come from discrimination of sound sequences with regular versus irregular rhythms (Bharucha & Pryor, 1986), from discrimination of sequences with simple ratios versus complex frequency ratios (Schellenberg & Trehub, 1994), and from categorical speech perception (Holt, 2005, 2006). Studies examining the time course of auditory discrimination suggest that temporal integration may be subserved by at least two auditory sensory memory systems, one lasting up to 300 ms and another lasting up to several seconds (Cowan, 1984). Thus, events occurring within these time frames may dramatically interact with each other during auditory processing.

Auditory stream segregation or "streaming" is a phenomenon that has been used as a model for how the auditory system segregates sound patterns arising from two or more distinct sources (e.g., a cocktail party situation; Cherry, 1953) and integrates the elements of the segregated patterns into perceptual objects or "streams" (Bregman, 1990; Bregman & Campbell, 1971; Moore & Gockel, 2002; Snyder & Alain, 2007b; Van Noorden, 1975). Streaming is often studied by repeatedly alternating a low tone (A) and a high tone (B), with every other B tone omitted and replaced by silence (−), taking the form ABA−ABA−. . . (Van Noorden, 1975). When the frequency difference between the A and B tones (Δf) is small, and the presentation rate is slow, listeners typically hear a single stream of

Joel S. Snyder, Department of Psychiatry, Veterans Affairs Boston Healthcare System/Harvard Medical School, and Department of Psychology, Harvard University; Olivia L. Carter and Erin E. Hannon, Department of Psychology, Harvard University; Suh-Kyung Lee, Speech and Hearing Bioscience and Technology Program, Massachusetts Institute of Technology, Cambridge, Massachusetts; Claude Alain, The Rotman Research Institute, Baycrest Centre for Geriatric Care, Toronto, Ontario, Canada.

Joel S. Snyder was supported by the Clinical Research Training Program "Clinical Research in Biological and Social Psychiatry" from the National Institutes of Health (T32 MH16259-28). Olivia L. Carter was supported by the National Health and Medical Research Council (Australia) C. J. Martin Fellowship (368525). This research was also supported by grants from the Canadian Institutes of Health Research and the Hearing Foundation of Canada awarded to Claude Alain.

Correspondence concerning this article should be addressed to Joel S. Snyder, who is now at the Department of Psychology, University of Nevada, Las Vegas, 4505 Maryland Parkway, Mail Stop 5030, Las Vegas, NV 89154-5030. E-mail: joel.snyder@unlv.edu

alternating high and low frequency tones with a galloping rhythm (i.e., ABA−ABA−). When Δf is suitably large, and the presentation rate is sufficiently fast, listeners hear two separate streams of tones, each with a constant frequency and a metronome rhythm (i.e., A−A−A−A− and B—B—). Importantly, streaming does not occur instantly, but instead listeners tend to hear one stream at the beginning of the trial that perceptually splits into two streams after several seconds (Anstis & Saida, 1985; Bregman, 1978). This so called "buildup" of streaming appears to be influenced by attention (Carlyon, Cusack, Foxton, & Robertson, 2001; Snyder, Alain, & Picton, 2006; for a review, see Snyder & Alain, 2007b) and intention (Pressnitzer & Hupé, 2006; Van Noorden, 1975). Experiments that have used longer trials have shown that following the buildup and the initial switch to streaming, perception of ABA− patterns shows bistability with temporal dynamics similar to perception of ambiguous visual stimuli (Pressnitzer & Hupé, 2006).

The presence of a relatively slow buildup process implies the existence of temporal integration over several seconds, suggesting that streaming may be highly influenced by temporal context. For example, Bregman (1978) used a stimulus pattern consisting of two consecutive tone pairs, with a silent interpattern interval (—), taking the form ABAB—ABAB—. Participants were asked to manually adjust the stimulus-onset asynchrony (SOA) between successive A and B tones until they heard two segregated streams. The size of the interpattern silence strongly influenced whether listeners could hear streaming, with longer silences (up to 4 s) resulting in the highest thresholds, and shorter silences (down to 0 s) resulting in the lowest thresholds for perceiving two streams. In other words, the temporal proximity of preceding patterns powerfully influenced whether streaming did or did not occur. In another study, Beauvois and Meddis (1997) measured perception of streaming for ABAB test patterns that were preceded by an induction sequence consisting of only the A tone but with the same SOA as the test pattern. As they increased the duration of the silent interval between the induction and test sequences up to several seconds, listeners were less likely to hear streaming. To test the stimulus generality of context effects on streaming, Rogers and Bregman (1993) used a test sequence consisting of a repeating ABA− pattern that was preceded by one of several induction patterns: (a) an isochronous series of tones identical in frequency to the A tones presented at the same rate as the ABA− pattern but with longer duration tones; (b) an isochronous series of tones identical in frequency, duration, and rate to the A tones; (c) an irregular series of tones identical in frequency, mean duration, and mean presentation rate to the A tones; and (d) continuous white noise. The first three (non-noise) induction sequences all enhanced perception of streaming in the test sequence compared with the induction sequence with noise. Induction sequences with overall rate similar to the A tones were slightly better at inducing streaming than the sequence with a similar rate to the ABA− pattern. Previous studies thus suggest that stream segregation is facilitated by previous exposure to one or both of the tones in the pattern.

A more recent study used the mismatch negativity wave from event-related brain potentials to examine neural correlates of context effects on streaming (Sussman & Steinschneider, 2006). The test sequence was a series of low (A) and high (B) tones arranged in an ambiguous repeating ABBB pattern that was possible to hear as one stream or two streams. Three different context sequences immediately preceded the test sequences with no break in the rhythm: (a) only the A tones from the test sequence, (b) the same pattern as the test sequence but with a smaller frequency separation, and (c) the exact same pattern as the test sequence. Occasionally, an A tone in the test sequence had a deviant intensity. If one of the context sequences enhanced streaming more so than the other context sequences, the mismatch negativity should have been larger for deviants in the following test sequence. The mismatch negativity to the deviant only occurred when the context sequence consisted of only the A tones, suggesting that this context enhanced segregation of the A and B tones in the test sequence (cf. Beauvois & Meddis, 1997; Rogers & Bregman, 1993). These results further suggest that the context sequence with a small frequency separation relative to the test sequence did not facilitate streaming. However, the lack of behavioral data to compare with mismatch negativity makes it difficult to evaluate how closely the results reflect perception of streaming.

In the current study, we seek to better understand the effect of preceding stimulus context on streaming. In particular, we used a range of Δfs between A and B tones to investigate how the size of the frequency difference on previous trials influenced perception of streaming on the current trial. On each trial, we presented a short sequence of ABA− patterns that could take on one of four Δf values. We measured how perception of streaming on each trial was influenced by Δf on the previous trials. To help rule out high-level effects as the source of context effects, we provided participants with a conflicting conscious strategy (Experiment 2). We also measured the strength of context effects as a function of the size of the silent interval between trials (Experiment 3). Characterizing the time course of context effects on streaming in this manner may implicate particular underlying mechanisms, such as auditory sensory memory (Cowan, 1984). Given the lack of prior data on the effect of previous Δf, it is equally conceivable that larger Δf on the preceding trial will result in (a) no effect on perception of streaming during the current trial, (b) more perception of streaming on the current trial, a *facilitative* context effect on streaming, similar to buildup time (Anstis & Saida, 1985; Bregman, 1978), or (c) less perception of streaming on the current trial, a *contrastive* context effect, which has not been previously observed in studies of streaming.

Experiment 1

In this experiment, we reanalyzed published behavioral data (Snyder & Alain, 2007a; Snyder et al., 2006) to test for effects of previous Δf on perception of streaming in the current trial. This data set was appropriate for this purpose because a large number of trials were presented to participants, thus yielding a sufficient number of examples of all combinations of previous Δf and current Δf.

Method

Participants. Ten adults (6 men and 4 women; age range = 23–38 years; mean age = 29.5 years) participated after giving written informed consent according to the guidelines of the Baycrest Centre for Geriatric Care and the University of Toronto. All participants were right-handed except 1, and all had normal pure-tone thresholds (<20 dB HL from 250 to 8000 Hz in both ears).

Materials and procedure. Stimuli were pure-tone patterns of alternating low (A) and high (B) tones, with every other B tone omitted and replaced with silence (−), taking the form ABA−ABA−. . .. Within each trial, the A tone frequency was always 500 Hz, and the B tone frequency was 500, 625, 750, or 1000 Hz. This corresponds to approximate Δf levels of 0, 4, 7, and 12 semitones. Tone duration was 20 ms, with 2.5-ms rise and fall times. The SOA was 100 ms between adjacent A and B tones within each ABA− cycle. The silent duration (−) was also 100 ms. These stimuli were generated by a Tucker Davis Technologies (Alachua, FL) RP-2 real-time processor (24-bit, 90 kHz bandwidth) that was controlled by a custom Matlab script on a Dell computer with a Pentium 4 processor. The analog outputs were fed into a Headphone driver (Tucker Davis Technologies HB-7), which were then transduced and presented binaurally through Sennheiser HD 265 headphones (Sennheiser Electronic Corporation, Old Lyme, CT) at about 85 dB SPL.

On each trial, participants were presented with 10.8 s of the ABA− pattern (27 ABA− repetitions). Within a block, 80 trials were presented in which Δf varied pseudorandomly from trial to trial (20 trials per Δf level). Participants were instructed to fixate a target on a screen in front of them and indicate at the end of the trial by pressing a button if they heard the pattern as one stream for the whole trial and another button if they heard the pattern split into two streams by the end of the trial. The experiment was self-paced, and the next trial began 2,000 ms after the response. Participants were instructed to focus on the rhythm as a cue (i.e., not galloping or galloping) to indicate whether the pattern was perceived as one stream or two streams. They were also instructed to let their perception take a natural time course rather than biasing themselves toward hearing the patterns in one way or another. Each participant performed four blocks for a total of 320 trials in each experimental condition (80 per Δf level). Prior to the experiment, participants completed eight practice trials with two examples of each Δf level. The experimental session lasted around 75 min.

Data analysis. For each current Δf level, we quantified the proportion of trials that participants heard as two streams as a function of the previous Δf levels occurring either at Lag 1 (for the trial immediately before the current trial), Lag 2 (for the trial two positions before the current trial), or Lag 3 (for the trial three positions before the current trial). Note that because Δf was randomly selected on each trial, the probability that a previous Δf at any particular value was the same across all lags.

We analyzed the proportion of trials in which participants reported hearing two streams using a three-way repeated measures analysis of variance (ANOVA) with current Δf (0, 4, 7, and 12 semitones), previous Δf (0, 4, 7, and 12 semitones), and lag (1, 2, 3) as factors. The degrees of freedom were adjusted with the Greenhouse–Geisser epsilon to correct for sphericity violations. All reported probability estimates were based on the reduced degrees of freedom, although the original degrees of freedom are reported.

Results and Discussion

The data revealed a main effect of current Δf with the likelihood of reporting hearing two streams increasing with Δf, $F(3, 27) = 88.58$, $p < .001$ (see Figure 1, upper panel). Figure 1 (upper panel) shows that larger Δf levels on the previous trial decreased the likelihood that participants reported streaming in the current trial, $F(3, 27) = 24.68$, $p < .001$, a contrastive effect of stimulus context (cf. Holt, 2005, 2006; Jones, Love, & Maddox, 2006).

An interaction between the current Δf and previous Δf, $F(9, 81) = 3.94$, $p < .025$, is consistent with the observation that previous Δf had the strongest influence on intermediate levels of current Δf. The effect of previous Δf was stronger for Lag 1 (immediately previous trial) compared with Lag 2 (see Figure 1, middle left) and Lag 3 (see Figure 1, lower left), as suggested by an interaction between previous Δf and lag, $F(3, 27) = 6.28$, $p < .001$. No other main effects or interactions were found.

A separate ANOVA was performed for each of the three lags to determine whether there was an influence of previous trials at each lag. For Lag 1, as the previous Δf increased, participants reported significantly less streaming in the current trial, $F(3, 27) = 24.04$, $p < .001$. Current Δf and previous Δf showed a significant interaction, $F(9, 81) = 4.37$, $p < .025$, with reduced effects of the previous Δf for extreme values of the current Δf (0 and 12 semitones). Lag 2 also showed a significant effect of the previous Δf on perception of streaming, $F(3, 27) = 14.31$, $p < .001$, but no interaction between current Δf and previous Δf, $F(9, 81) = 1.91$, *ns*. Finally, Lag 3 showed no effect of previous Δf, $F(3, 27) = 1.90$, *ns*, and no interaction between current Δf and previous Δf, $F(9, 81) = 1.04$, *ns*, suggesting that the effect of context faded away after about two trials.

In summary, perception of streaming was most likely when the Δf between the A and B tones in an ABA− pattern was large (Bregman, 1990; Moore & Gockel, 2002), but larger Δf on previous trials decreased perception of streaming in the current trial. The effect of previous Δf was observed for Lag 2 but not for Lag 3, suggesting that the influence of previous trials lasted for at least 12.8 s (i.e., the time between the previous response and the end of the current trial). These results suggest that perception of streaming is affected by previous trials and that this effect lasts for a relatively long time, corresponding to the approximate duration of auditory sensory memory (Cowan, 1984). However, the observed effects could be explained by a conscious strategy, whereby participants tried to vary their responses from one trial to the next by picking the opposite response. To assess the possibility of such higher level biases, in Experiment 2 we examined whether listeners could purposefully change their perception and diminish the context effect by trying to hear a pattern as either segregated or as integrated on every trial.

Experiment 2

To determine the robustness of the observed effect of context from Experiment 1 in the face of a competing high-level influence on perception of streaming, we asked participants for one half of Experiment 2 to try and hear the ABA− patterns as one stream and for the other half of the experiment to try to hear two streams. Previous studies have shown that participants were able to substantially manipulate their perception of streaming, especially for intermediate levels of Δf (Pressnitzer & Hupé, 2006; Van Noorden, 1975). We expected that if the context effect observed in Experiment 1 was a true perceptual effect that occurred without participants' awareness, it would remain even though participants were intentionally biasing their perception. If on the other hand, the context effect was due to participants intentionally varying their responses from one trial to the next, providing a specific strategy to always hear the patterns as one way or another should override the incompatible strategy of varying responses.

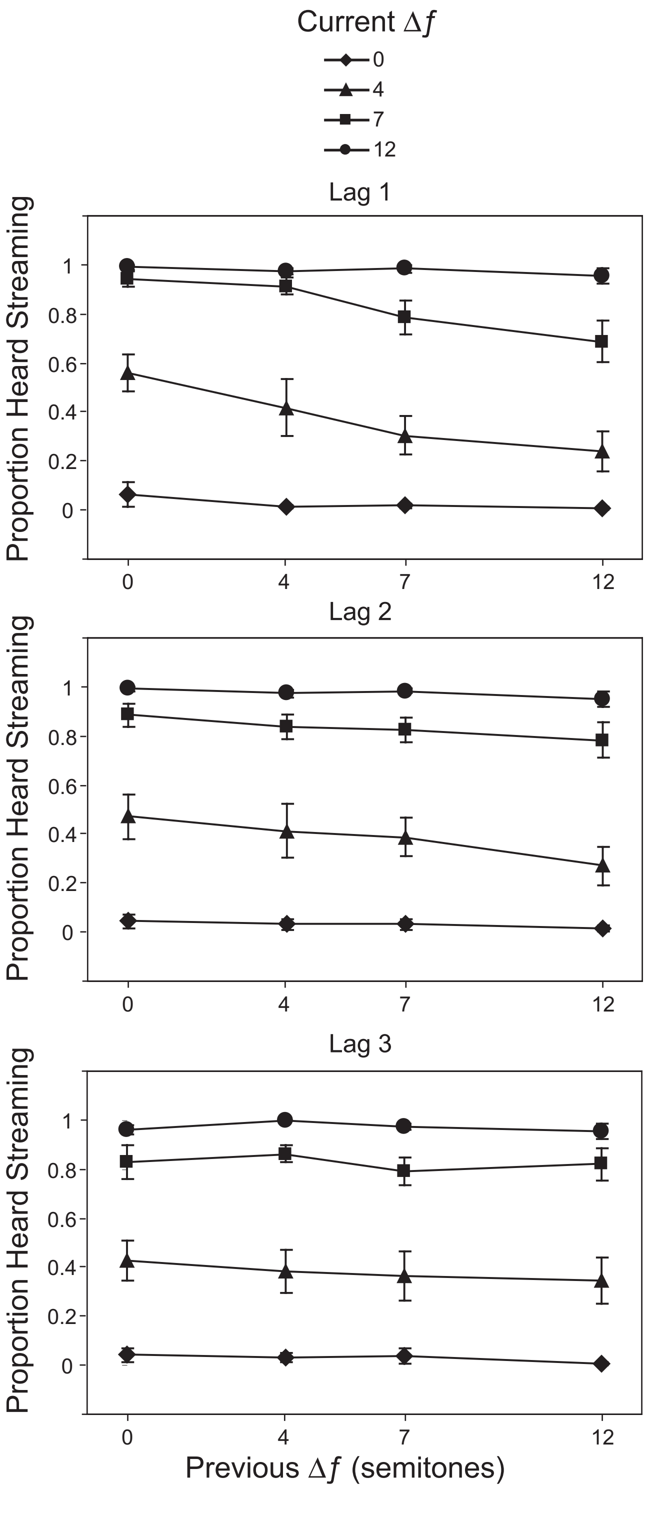

Figure 1. Proportion of trials heard as streaming in Experiment 1, showing the effect of the previous and current frequency separation between low (A) and high (B) tones (Δf = 0, 4, 7, or 12 semitones). Separate lines are for different values of Δf on the current trial. The influences of the immediately preceding trial (Lag 1, top), two trials before the current one (Lag 2, middle), and three trials before the current one (Lag 3, bottom) are shown separately. Note that a negative slope as a function of previous Δf indicates a contrastive effect of context because this is opposite of the effect of current Δf. Error bars represent the standard error of the mean.

Method

Participants. Ten adults (5 men and 5 women; age range = 21–37 years; mean age = 29.3 years) participated after giving written informed consent according to the guidelines of the Baycrest Centre for Geriatric Care and the University of Toronto. All participants were right-handed except 1, and all had normal pure-tone thresholds (<20 dB HL) at frequencies from 250 to 8000 Hz in both ears. Joel S. Snyder (the first author) participated in this experiment.

Materials and procedure. The same materials and procedure were used as in Experiment 1, with the following changes. Instead of letting perception take a natural course as in Experiment 1, half of the participants were instructed to try hearing the ABA− patterns as one stream (*integrate* condition) for the first two blocks of trials and as two streams (*segregate* condition) for the last two blocks of trials. The other half of participants were in the segregate condition for the first two blocks and the integrate condition for the last two blocks.

Data analysis. The analysis of behavioral responses was the same as in Experiment 1, with the exception that intention (integrate vs. segregate) was added as a factor in the ANOVAs.

Results and Discussion

Figure 2 shows the proportion of trials heard as streaming for each level of Δf as a function of the previous Δf for Lag 1, Lag 2, and Lag 3, separately for the integrate and segregate conditions. Participants reported being able to modulate their perception, as shown by a greater amount of streaming reported in the segregate condition than in the integrate condition, especially for intermediate levels of Δf. This was confirmed by a main effect of intention on streaming, $F(3, 27) = 24.99$, $p < .001$, and an interaction between intention and current Δf, $F(3, 27) = 10.81$, $p < .001$. There were no other interactions between intention and other factors, suggesting that the context effect observed in Experiment 1 was not due to participants intentionally varying their responses from trial to trial. As in Experiment 1, participants reported more streaming with larger current Δf, $F(3, 27) = 85.22$, $p < .001$, and, despite the intention to control their perception, less streaming occurred when previous Δf was larger, $F(3, 27) = 13.25$, $p < .001$. As before, the effect of the previous trial diminished for longer lags, $F(3, 27) = 3.46$, $p < .05$, and the effect of previous Δf was stronger for intermediate values of current Δf, $F(9, 81) = 4.00$, $p < .025$.

For Lag 1, there was a significant influence of the previous Δf on streaming, $F(3, 27) = 18.07$, $p < .001$, and the effect was larger at intermediate levels of Δf as indicated by an interaction between

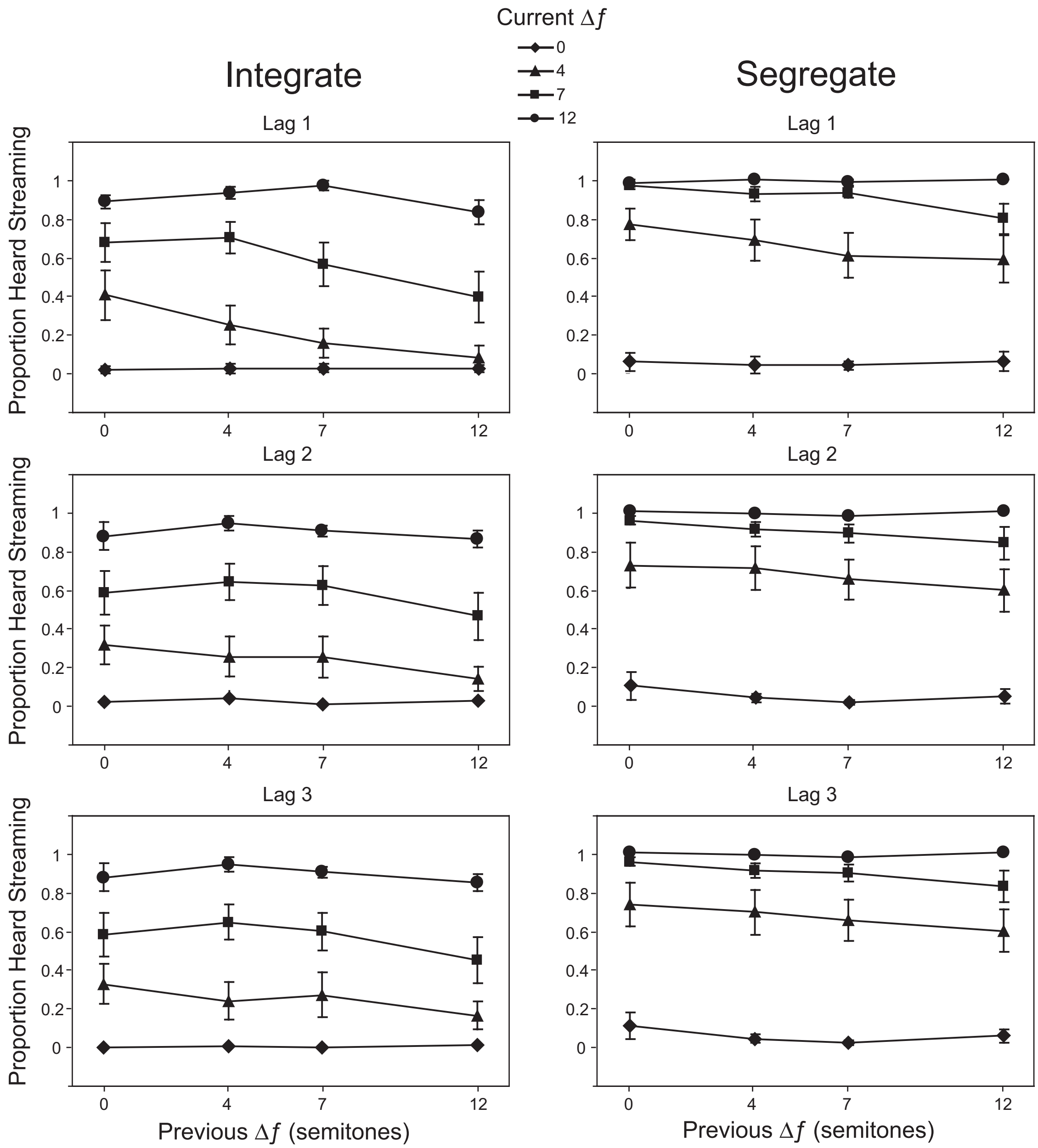

Figure 2. Proportion of trials heard as streaming in Experiment 2, showing the effect of the previous and current Δf. The two columns show the influence of the previous trials when participants intended to perceive one stream of tones (Integrate, left column) and when participants intended to hear two streams of tones (Segregate, right column). Error bars represent the standard error of the mean.

current Δf and previous Δf, $F(9, 81) = 6.21$, $p < .005$. As with Lag 1, for Lag 2 and Lag 3, the larger the previous Δf, the less participants reported streaming, $F(3, 27) = 8.62$ and 7.72, $p < .005$, respectively. For Lag 2 and Lag 3, however, the interaction between the previous and current Δf was no longer significant, $F(3, 27) = 2.14$ and 2.35, *ns*, respectively. Thus, the influence of context in Experiment 2 appears to have lasted longer than in Experiment 1. This could be due to the increased task demand of intentionally hearing the stimuli as one stream or two streams or it could be due to using a different group of participants.

The results of Experiment 2 replicated Experiment 1 by showing that the likelihood of hearing streaming is modulated by prior context, with larger previous Δf decreasing perception of streaming on the current trial. We extended this contrastive effect of context by demonstrating its robustness even when participants intentionally modulated their perception by trying to hear one stream or two streams. This diminishes, but does not eliminate, the plausibility of nonperceptual explanations of the observed context effects, such as a conscious strategy of giving varied responses. Therefore, it is possible that the two strategies could have summated. It is also possible that some sort of nonconscious response bias could have resulted in the context effect, a possibility that we address in the next experiment.

Experiment 3

In Experiments 1 and 2, the influence of previous Δf on streaming in the current trial decreased with time between trials. The purpose of Experiment 3 was to investigate the time course of the context effect in a more controlled manner to determine whether it decays over a matter of a few seconds, which would be consistent with auditory sensory memory and/or sensory adaptation processes. We therefore manipulated the intertrial interval within blocks of trials. In Experiments 1 and 2, participants gave a single response per trial, which introduced uncontrolled variability in the duration between trials. Experiment 3 therefore continuously measured participants' responses throughout each trial, which allowed us to precisely control the duration of intertrial intervals and to assess the time course of the context effect.

Method

Participants. Twenty adults (8 men and 12 women; age range = 22–49 years; mean age = 28.1 years) from the Harvard University community participated after giving written informed consent according to the guidelines of the Faculty of Arts and Sciences at Harvard University. All participants were right-handed except 3 who were left-handed and 1 who was ambidextrous. All reported having normal hearing. Joel S. Snyder, Olivia L. Carter, Suh-Kyung Lee, and Erin E. Hannon (the first four authors) participated in this experiment.

Materials and procedure. As in Experiments 1 and 2, stimuli were pure-tone ABA− patterns. Stimuli were generated and behavioral responses were collected by a custom Matlab script that used functions from the Psychtoolbox (Brainard, 1997), running on an IBM PC Pentium 4 computer with a SoundMAX Integrated Digital Audio sound card. The sounds were presented binaurally through Koss (Milwaukee, WI) UR-30 closed ear headphones at 65 dB SPL. Within each trial, the A tone frequency was always 500 Hz, and the B tone frequency was 500, 600, 700, or 1000 Hz. This corresponds to approximate Δf levels of 0, 3, 6, and 12 semitones. Tone duration was 50 ms, with 10-ms rise and fall times. The SOA was 120 ms between adjacent A and B tones within each ABA− cycle. The silent duration between ABA triplets was also 120 ms. The frequency and duration values of the stimuli in this experiment were slightly different than in Experiments 1 and 2 but yielded similar overall proportions of hearing one stream or two streams.

On each trial, participants were presented with 12.96 s of the ABA− pattern (27 ABA− repetitions). Each participant was presented with trials in three sets of five blocks, with each set presenting one of three silent durations between trials. These silent intertrial intervals were 1.44, 4.32, or 5.76 s (i.e., 3, 9, or 12 ABA− cycles in duration), and sets were presented in a random order to each participant. For each set, the five blocks of 16 trials were presented with pseudorandom orders of Δf levels, with the constraint that all but 1 of the 16 serial combinations of four previous Δfs and four current Δfs would be presented exactly once in each block. It was not possible to present all 16 serial combinations within a block of 16 trials because there were only 15 trials presented with a preceding trial (i.e., the first trial did not have a preceding trial). Between each block of trials, participants could take a short break for as long as they wanted before beginning the next block.

Participants were instructed to fixate a cross on the screen and listen to the ABA− patterns. As soon as possible after the beginning of each trial they were asked to press the down-arrow key or the right-arrow key to indicate whether they perceived one stream or two streams, respectively. Participants were further instructed to hold down the button for as long as they experienced the corresponding perception, and if the perception switched at any point during the trial, participants were instructed to switch buttons accordingly. As in Experiment 1, participants were encouraged to let their perception take a natural course and not to bias their perception in favor of one stream or two streams. The program recorded which button was being pressed synchronously with the A tones (i.e., once every 240 ms), resulting in 54 data points per trial. The experimental session lasted around 75 min.

Data analysis. We quantified the proportion of trials in which participants perceived two streams at each sampled time point over the course of the trial, separately for each combination of current Δf, previous Δf, and intertrial interval. For each of these conditions, we took the mean value across the second half of the trial (i.e., Samples 28–54) for each participant as a measure of steady-state perception of streaming following the buildup. The steady-state values were entered in a three-way repeated measures ANOVA with current Δf (0, 3, 6, and 12 semitones), previous Δf (0, 3, 6, and 12 semitones), and intertrial interval (1.44, 4.32, and 5.76 s) as factors. As in Experiments 1 and 2, the degrees of freedom were adjusted with the Greenhouse–Geisser epsilon, and all reported probability estimates were based on the reduced degrees of freedom.

Results and Discussion

Figure 3 shows the time course of streaming averaged across all participants. Consistent with the results of Experiments 1 and 2, participants reported hearing streaming most often when the current Δf was larger, $F(3, 57) = 49.20$, $p < .001$, and when the previous Δf was smaller, $F(3, 57) = 39.71$, $p < .001$, with a larger effect of previous Δf for intermediate values of current Δf, $F(9, 171) = 9.27$, $p < .025$. Increasing the intertrial interval diminished the effect of previous Δf on perception of streaming in the current trial, as indicated by an interaction between the previous Δf and intertrial interval, $F(6, 114) = 2.52$, $p < .05$, suggesting that the context effect begins decaying in the first few seconds after the previous trial is over. However, even for the longest intertrial interval, the effect of previous Δf remained, consistent with the Lag 2 and Lag 3 context effects in Experiments 1 and 2. No other main effects or interactions were significant. These results suggest that the observed context effect decays over the course of several seconds, as with facilitative context effects on streaming (Beauvois & Meddis, 1997; Bregman, 1978).

To further support the idea that the contrastive context effect we observed is an effect on sensory or perceptual processing, rather than some form of response bias, we reanalyzed the data according to perception (one stream or two streams) at the end of the previous trial for trials in which the current and previous Δf were

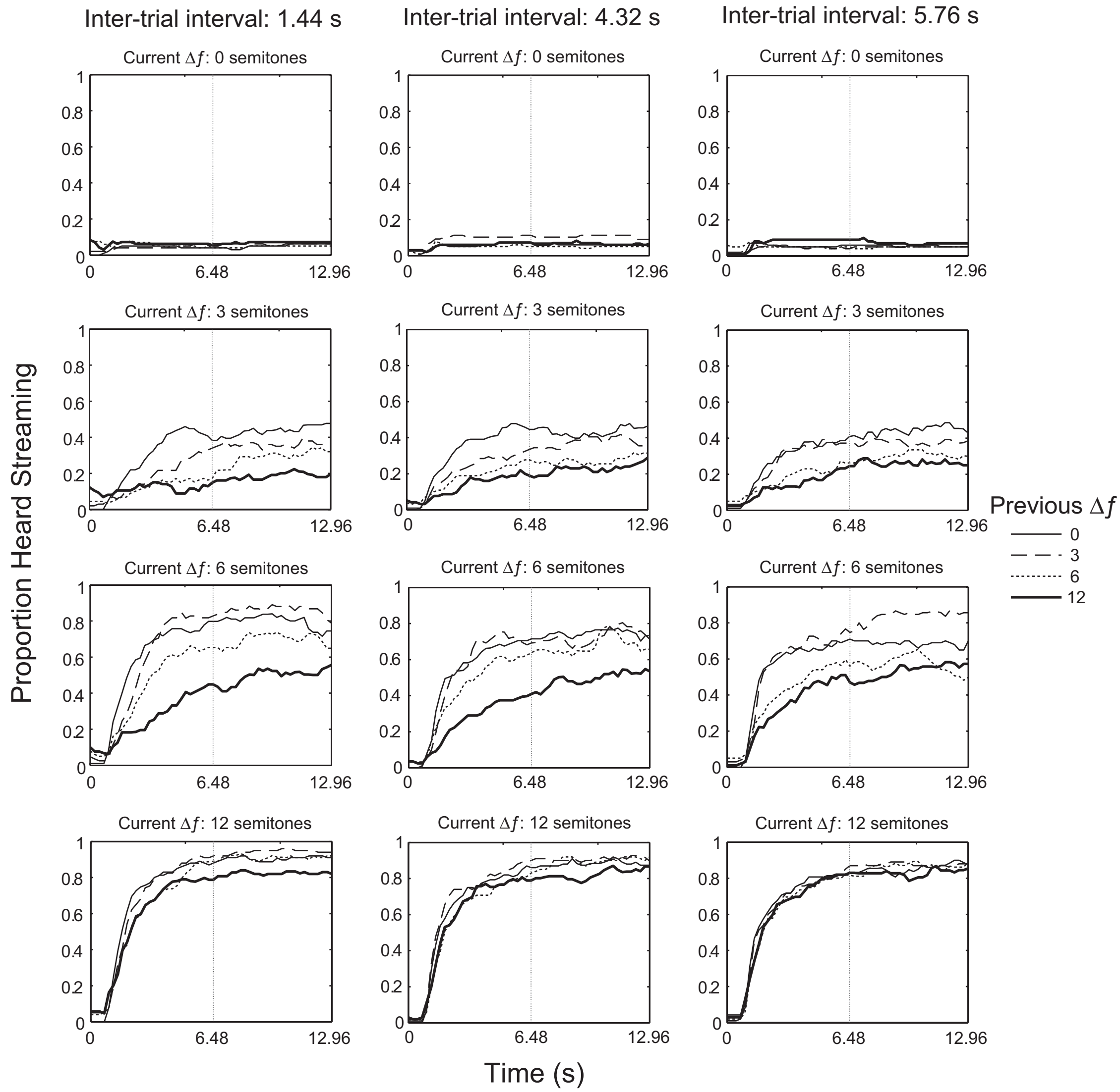

Figure 3. Proportion of trials heard as streaming over the time course of the trial in Experiment 3, showing the effect of the previous Δf (0, 3, 6, or 12 semitones) on perception of streaming in the current trial. Separate lines are for different values of Δf in the previous trial. The three columns show the influence of the previous trials for three different intertrial intervals (1.44, 4.32, and 5.76 s). Vertical dotted lines mark the middle point of the trial to indicate the beginning of the averaging window used for data analysis.

the same (regardless of the intertrial interval). If the effect of previous percept differs from that of previous Δf, then one may conclude that the effect of previous Δf does not reflect response bias. We performed this analysis only for the conditions in which the Δf was ambiguous (three and six semitones) so that the majority of participants would have examples of both previous percepts. Figure 4 shows the time course of streaming averaged across all participants that had at least one trial of a given previous percept/current Δf combination. In contrast to the effect of larger previous Δf on perception of streaming, perceiving two streams at the end of the previous trial did not cause less streaming during the current trial. Instead, it appears that perceiving two streams on the previous trial actually caused more streaming during the current trial, although we did not test this effect statistically because of the unequal numbers of participants for each combination of previous percept/current Δf. Nevertheless, this analysis is inconsistent with the effect of previous Δf as solely reflecting response bias.

General Discussion

Three experiments were conducted to investigate the effect of stimulus context on auditory stream segregation. Participants exhibited the well-known tendencies to increasingly perceive two streams as the difference between the A and the B tones increased (Bregman & Campbell, 1971; Van Noorden, 1975) and as the number of ABA− repetitions increased (i.e., buildup; Anstis & Saida, 1985; Bregman, 1978). In contrast to the effects of current

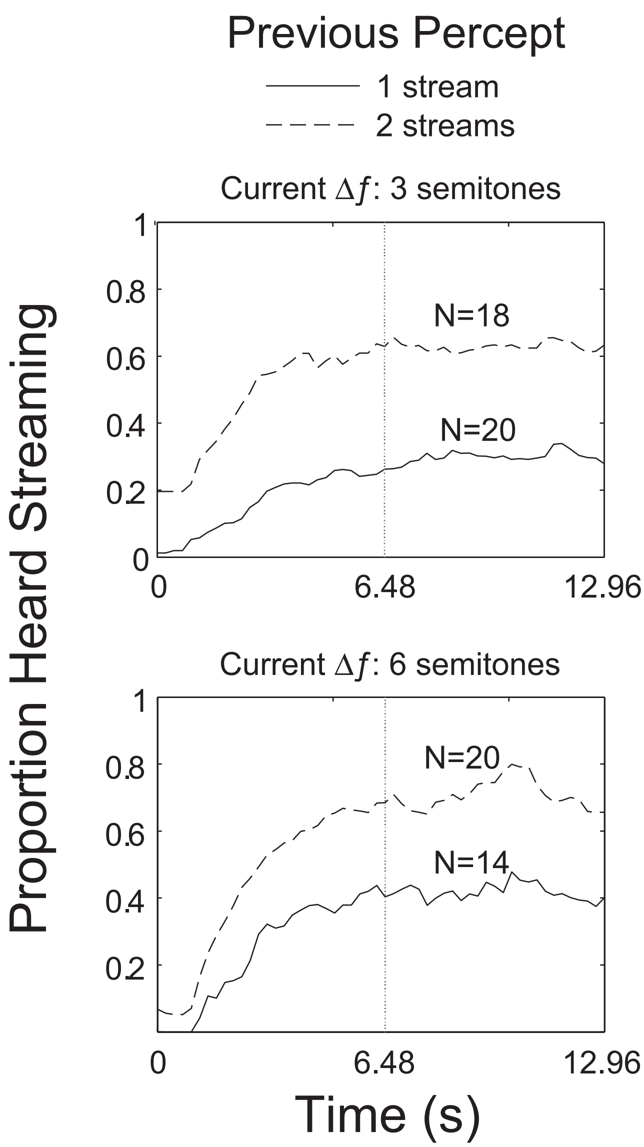

Figure 4. Proportion of trials heard as streaming over the time course of the trial in Experiment 3, showing the effect of the previous percept (one stream or two streams) on perception of streaming in the current trial. Separate lines are for different percepts in the previous trial. Only the two conditions with ambiguous current Δf (three and six semitones) are shown. Note that unequal numbers of participants are represented in each of the average time courses because not all participants had at least one trial of each previous percept for each current Δf. Vertical dotted lines mark the middle point of the trial to indicate the beginning of the averaging window used for data analysis.

Δf and buildup, participants were more likely to perceive two streams in the current trial when the previous trial had a smaller Δf. This contrastive context effect lasted for many seconds, as demonstrated by the influence of one or two trials before the previous one on the current trial (Experiments 1 and 2), but showed a decay beginning in the first few seconds after the end of the previous trial (Experiment 3). The contrastive nature of the context effect found in the current study is similar to findings from effects of context on categorical perception of speech sounds. For example, Holt (2005, 2006; also see Jones et al., 2006) showed that perception of "ga" versus "da" is influenced by prior presentation of tone sequences in a contrastive manner such that high-frequency tones resulted in more perception of "ga" despite the fact that "ga" had lower frequency formants than "da." The fact that contrastive context effects occur in both speech perception and stream segregation point to a general phenomenon that may arise from similar neurocomputational processes operating on auditory information.

Effects of Context on Streaming: Comparison With Previous Studies

Previous behavioral studies of context (Beauvois & Meddis, 1997; Bregman, 1978; Rogers & Bregman, 1993) did not address whether context sequences with smaller or larger Δf caused more streaming. However, an event-related brain potential study found a mismatch negativity to a deviant tone only when the induction sequence consisted of the A tones from a repeating ABBB test sequence and not when the context was an ABBB sequence with a smaller Δf than the test sequence (Sussman & Steinschneider, 2006). Because responses to deviant tones are thought to arise when streaming occurs, the authors concluded that no streaming occurred when test sequences were preceded by a smaller Δf. Because no behavioral responses were collected to compare with the mismatch negativity, the present findings cast doubt on the extent to which the mismatch negativity observed by Sussman and Steinschneider (2006) truly reflected an effect of context on perception of streaming.

An alternative explanation of the finding that larger previous Δf reduced streaming on the current trial is that participants were intentionally varying their responses from trial to trial. This explanation is highly unlikely given that sorting the data from Experiment 3 according to prior percept when the current and previous Δf were the same did not reproduce the contrastive context effect. Instead, the trend was in the opposite direction, with more perception of two streams on the previous trial leading to more perception of two streams on the current trial. Thus, the Δf-related contrastive context effect observed in the present study is likely to be a genuine sensory or perceptual effect and is not due to response bias.

The current study further showed that the effect of previous trials (Lag 1) could also occur for the trial before the previous one (Lag 2, Experiments 1 and 2) and for two trials before the previous one (Lag 3, Experiment 2), although these longer range context effects were not as strong as the Lag 1 effect. These findings suggest that the influence of context lasts for longer than 25.6 s (i.e., the duration of two trials plus two intertrial intervals). Varying the intertrial interval from 1.44 to 5.76 s decreased the size of the context effect (Experiment 3), suggesting that although the context effect lasts for tens of seconds, it begins decaying after just a few seconds. This result is consistent with a previous finding that the biasing effect of context showed the steepest decline for delays of 0–0.7 s, with more gradual declines occurring in the next few seconds (Bregman, 1978; also see Beauvois & Meddis, 1997). The time course of the decay of stream biasing is similar to the time course of streaming buildup (Anstis & Saida, 1985; Bregman, 1978), suggesting that perception of streaming is associated with relatively long time constants of integration. The relatively long temporal integration periods associated with streaming are also similar to the duration of auditory sensory memory (Cowan, 1984) and to electrophysiological correlates of auditory temporal integration (Lu, Williamson, & Kaufman, 1992; Näätänen & Winkler, 1999). Such a long temporal integration of information may be particularly important in the auditory system because acoustic patterns, such as speech and music, evolve over many seconds.

Despite the similar temporal dynamics of context effects and buildup of streaming, it is not clear whether these two processes rely on similar neural mechanisms, thus raising a number of interesting empirical questions. One clear difference between the two phenomena is that buildup can occur when the context is a single repeating tone (Beauvois & Meddis, 1997) or when it consists of alternating tones (Bregman, 1978), whereas the contrastive context effect only occurs when the context consists of alternating tones. Although there is evidence that buildup processes depend on actively attending to the ABA− sequences (Carlyon et al., 2001; Snyder et al., 2006; but see Macken, Tremblay, Houghton, Nicholls, & Jones, 2003), it is not known whether this is also true of context effects. Similarly, although there is neurophysiological and computational evidence for neural adaptation effects in auditory cortex underlying the buildup of streaming (McCabe & Denham, 1997; Micheyl, Tian, Carlyon, & Rauschecker, 2005), it is not clear whether central (McCabe & Denham, 1997) or peripheral (Beauvois & Meddis, 1996) processes underlie effects of context on streaming. Thus, future studies should manipulate attention and measure neurophysiological activity while concurrently measuring context effects on perception of streaming to address the extent to which buildup and context effects rely on similar neurocomputational mechanisms.

Neural Adaptation as a Possible Mechanism for Context Effects

A slow form of neural adaptation was recently shown to predict the buildup of streaming (Micheyl et al., 2005). Monkeys were trained to listen attentively to trials of 10-s ABA− patterns while single-unit responses from neurons tuned to the A tone frequency were recorded in primary auditory cortex. The neural response to the B tone decreased with larger Δf between the A and B tones, whereas the response to the A tones did not change with increasing Δf. Over the course of the trial, neural responses to both the A and B tones declined. This decrease in responsiveness over time, along with the differential effect of Δf on responses to the A and B tones, provided sufficient information to predict whether perception of one stream or two streams was occurring in a set of behavioral data obtained from human listeners. Slow neural adaptation occurring in response to long patterns of repeating ABA− sequences could also underlie the context effects observed in the current study.

Given the similar time course of context effects on perception of streaming, slow neural adaptation occurring in response to long tone patterns (Micheyl et al., 2005; also see Ulanovsky, Las, Farkas, & Nelken, 2004; Ulanovsky, Las, & Nelken, 2003) could potentially underlie the context effect observed in the current study and in other studies (Beauvois & Meddis, 1997; Bregman, 1978). Neural adaptation to specific tone frequencies is especially plausible for facilitative context effects, such as buildup, given that they can occur even when the context is a repeating tone of a single frequency (Beauvois & Meddis, 1997; Rogers & Bregman, 1993). However, it is more difficult to explain how adaptation to single frequencies could explain the contrastive context effect observed here that results from varying Δf. Instead, it is possible that the Δf-dependent context effect depends on adaptation of frequency-shift detectors. Specifically, adaptation of neurons tuned to large frequency shifts would result in a greater proportion of responses from neurons tuned to small frequency shifts, resulting in more perception of one stream; conversely, adaptation of neurons tuned to small frequency shifts would result in a greater proportion of responses from neurons tuned to large frequency shifts, resulting in more perception of two streams. The notion of frequency-shift detectors has been proposed previously as an explanation for streaming (Anstis & Saida, 1985; Van Noorden, 1975) and melodic interval perception (Demany & Ramos, 2005). Future research should test more directly for adaptation of frequency shift detectors as an explanation for the context effects we observed. The possibility that stream segregation engages auditory processes that are sensitive to relative frequency (i.e., frequency shifts) in addition to absolute frequency (i.e., tonotopic proximity) would have important implications for understanding the different types of auditory coding that support streaming (for a review, see Snyder & Alain, 2007b) and auditory perception more generally.

Summary

In three experiments, we found robust effects of context on auditory stream segregation. Specifically, stimulus patterns containing a large frequency separation on the previous trial resulted in reduced perception of streaming in the current trial. A similar effect occurred for the trial before the previous one and the trial before that, suggesting long-lasting effects of context on perceptual organization. However, the effect of context began decaying after several seconds, as demonstrated by manipulating the time between consecutive trials, suggesting that the observed context effect has a similar time course to long auditory sensory memory (Cowan, 1984) and neural adaptation in auditory cortex (Micheyl et al., 2005; Ulanovsky et al., 2003, 2004).

References

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

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

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

Bharucha, J. J., & Pryor, J. H. (1986). Disrupting the isochrony underlying rhythm: An asymmetry in discrimination. *Perception & Psychophysics, 40,* 137–141.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436.

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

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.

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.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America, 25,* 975–979.

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin, 96,* 341–370.

Demany, L., & Ramos, C. (2005). On the binding of successive sounds: Perceiving shifts in nonperceived pitches. *The Journal of the Acoustical Society of America, 117,* 833–841.

Fecteau, J. H., & Munoz, D. P. (2003). Exploring the consequences of the previous trial. *Nature Reviews Neuroscience, 4,* 435–443.

Gilden, D. L. (2001). Cognitive emissions of 1/f noise. *Psychological Review, 108,* 33–56.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences, 10,* 14–23.

Holt, L. L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychological Science, 16,* 305–312.

Holt, L. L. (2006). The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization. *The Journal of the Acoustical Society of America, 120,* 2801–2817.

Jones, M., Love, B. C., & Maddox, W. T. (2006). Recency effects as a window to generalization: Separating decisional and perceptual sequential effects in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 316–332.

Lu, Z. L., Williamson, S. J., & Kaufman, L. (1992, December 4). Behavioral lifetime of human auditory sensory memory predicted by physiological measures. *Science, 258,* 1668–1670.

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.

McCabe, S. L., & Denham, M. J. (1997). A model of auditory streaming. *The Journal of the Acoustical Society of America, 101,* 1611–1621.

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.

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

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin, 125,* 826–859.

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

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

Schellenberg, E. G., & Trehub, S. E. (1994). Frequency ratios and the discrimination of pure tone sequences. *Perception & Psychophysics, 56,* 472–478.

Snyder, J. S., & Alain, C. (2007a). Sequential auditory scene analysis is preserved in normal aging adults. *Cerebral Cortex, 17,* 501–512.

Snyder, J. S., & Alain C. (2007b). Toward a neurophysiological theory of auditory stream segregation. *Psychological Bulletin, 133,* 780–799.

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.

Sussman, E., & Steinschneider, M. (2006). Neurophysiological evidence for context-dependent encoding of sensory input in human auditory cortex. *Brain Research, 1075,* 165–174.

Ulanovsky, N., Las, L., Farkas, D., & Nelken, I. (2004). Multiple time scales of adaptation in auditory cortex neurons. *Journal of Neuroscience, 24,* 10440–10453.

Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature Neuroscience, 6,* 391–398.

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

Received November 22, 2006
Revision received September 14, 2007
Accepted October 14, 2007 ■