

Objective and Subjective Psychophysical Measures of Auditory Stream Integration and Segregation

CHRISTOPHE MICHEYL¹ AND ANDREW J. OXENHAM¹

¹*Department of Psychology, University of Minnesota, N640b Elliott Hall, 75 East River Road, Minneapolis, MN 55455-0344, USA*

Received: 20 November 2009; Accepted: 30 June 2010; Online publication: 24 July 2010

ABSTRACT

The perceptual organization of sound sequences into auditory streams involves the integration of sounds into one stream and the segregation of sounds into separate streams. “Objective” psychophysical measures of auditory streaming can be obtained using behavioral tasks where performance is facilitated by segregation and hampered by integration, or vice versa. Traditionally, these two types of tasks have been tested in separate studies involving different listeners, procedures, and stimuli. Here, we tested subjects in two complementary temporal-gap discrimination tasks involving similar stimuli and procedures. One task was designed so that performance in it would be facilitated by perceptual integration; the other, so that performance would be facilitated by perceptual segregation. Thresholds were measured in both tasks under a wide range of conditions produced by varying three stimulus parameters known to influence stream formation: frequency separation, tone-presentation rate, and sequence length. In addition to these performance-based measures, subjective judgments of perceived segregation were collected in the same listeners under corresponding stimulus conditions. The patterns of results obtained in the two temporal-discrimination tasks, and the relationships between thresholds and perceived-segregation judgments, were mostly consistent with the hypothesis that stream segregation helped performance in one task

and impaired performance in the other task. The tasks and stimuli described here may prove useful in future behavioral or neurophysiological experiments, which seek to manipulate and measure neural correlates of auditory streaming while minimizing differences between the physical stimuli.

Keywords: auditory streaming, segregation, integration, duration discrimination

INTRODUCTION

Although acoustic stimuli such as speech and music often consist of series of sounds (syllables or notes), we usually do not usually experience these sounds as unrelated auditory events, but rather as coherent “streams,” which can be followed over time as a single entity. The perceptual organization of sounds into streams, known as “auditory streaming,” is an important aspect of “auditory scene analysis” (Bregman 1990). The phenomenon can be demonstrated using sequences of tones organized temporally in a repeating ABAB... or ABA-ABA-... pattern, where A and B indicate tones of different frequencies, and the dash indicates a silent gap (Miller and Heise 1950; van Noorden 1975)—audio examples can found at <http://www.tc.umn.edu/~cmicheyl/demos.html>. When the frequency difference between the A and B tones is relatively small (e.g., a semitone, or one twelfth of an octave), most listeners hear the sequence as a coherent stream, a perceptual state that is sometimes referred to as “stream integration.” However, if the A-B frequency separation is large (e.g., an octave or more), listeners usually report hearing “two

Correspondence to: Christophe Micheyl · Department of Psychology · University of Minnesota · N640b Elliott Hall, 75 East River Road, Minneapolis, MN 55455-0344, USA. Telephone: +1-612-6263291; fax: +1-612-6263359; email: cmicheyl@umn.edu

streams”, one at a low pitch, the other at a high pitch, a perceptual state that is commonly referred to as stream segregation. Listeners can then selectively attend to either stream. Whether one or “two streams” are heard depends on factors other than just frequency separation. The tone repetition rate, or the inter-tone interval, and the number of tones in the sequence also play an important role. In general, faster presentation rates, or shorter tone intervals, and longer sequence lengths promote segregation (van Noorden 1975; Bregman 1978; Bregman et al. 2000). Auditory streaming has been demonstrated using a variety of sounds, including pure tones, harmonic complex tones (Cusack and Roberts 1999; Vliegen et al. 1999; Vliegen and Oxenham 1999; Grimault et al. 2000, 2001; Roberts et al. 2002, 2008), synthetic vowels (Gaudrain et al. 2007, 2008), and noises (Bregman et al. 2001; Grimault et al. 2002), with streaming depending on differences in fundamental frequency (F0), spectral content, or temporal envelope modulation rate.

The phenomenon of auditory streaming has inspired a large number of psychophysical studies over the past 50 years (reviewed in Darwin and Carlyon 1995a, b; Moore and Gockel 2002; Carlyon and Gockel 2008). More recently, it has started to attract the interest of both animal behaviorists and neuroscientists (for reviews, see Carlyon 2004; Micheyl et al. 2007b; Snyder and Alain 2007; Bee and Micheyl 2008; Fay 2008; Shamma and Micheyl 2010). In this context, it is desirable for experimenters to have at their disposal measures of auditory streaming that do not rely on reports of perceived segregation (such as “one stream” *versus* “two stream” judgments).

One approach to measuring auditory streaming percepts without relying on reports of perceived segregation involves having a listener perform a perceptual task, the performance of which is influenced by stream segregation. Research in psychoacoustics has already led to the identification of such tasks. For instance, several studies have shown that listeners are poorer at identifying the temporal order of sounds, or detecting changes in their relative timing, if these sounds fall into separate streams than if they fall into the same stream (van Noorden 1975; Vliegen et al. 1999; Roberts et al. 2002, 2008; Micheyl and Oxenham 2010). In such tasks, therefore, stream segregation impedes performance. A recent neurophysiological study of auditory streaming has taken advantage of this (Elhilali et al. 2009). Conversely, several studies have found that listeners can more accurately recognize a familiar melody, that is temporally interleaved with another melody, if the two melodies form separate streams (Dowling 1973; Hartmann and Johnson 1991; Vliegen and Oxenham

1999; Bey and McAdams 2002, 2003). Presumably, this is due to listeners being able to attend selectively to individual streams. Thus, in this type of task, stream segregation improves performance. Neurophysiological studies of auditory streaming in humans have taken advantage of this (e.g., Sussman et al. 1999). Other examples of tasks in which stream segregation appears to facilitate performance include the discrimination of pitch sequences in the presence of temporal flankers (Micheyl and Carlyon 1998; Gockel et al. 1999; Micheyl et al. 2005b), tone detection in the presence of simultaneous multi-tone maskers (Kidd et al. 1994, 2002; Durlach et al. 2003; Kidd et al. 2003; Oxenham et al. 2003; Huang and Richards 2006; Micheyl et al. 2007a), the detection of amplitude modulation in the presence of interfering modulation in a remote frequency region (Oxenham and Dau 2001), and binaural interference (Best et al. 2008).

Tasks in which stream segregation either improves or impedes performance have almost invariably been tested in separate studies, involving different listeners, different stimuli, and different experimental procedures. In fact, we are aware of only one study in which performance was measured in the same listeners using both types of tasks, with comparable stimuli (Micheyl et al. 2005b). The pattern of results obtained in that study were generally consistent with the hypothesis that performance in the two tasks, which involved the detection of amplitude-modulated tones temporally interspersed among steady or amplitude-modulated tones, was related to auditory streaming. Unfortunately, the conclusions of that study were limited by the fact that the performance measures were not compared with direct measures of auditory streaming, where subjects are asked whether the sequence was heard as one or “two streams”. As such, it was not possible to perform a direct comparison of task performance and perception.

Here, we describe two temporal-perception tasks, which were designed in such a way that stream segregation should facilitate performance in one task and hamper performance in the other task. These tasks use very similar stimuli. Thresholds in these two tasks were measured in the same listeners under different stimulus conditions, which were produced by varying three parameters that have been shown in previous studies to influence the perceptual organization of tone sequences: frequency separation, tone-presentation rate, and sequence length. In addition to these objective psychophysical measures, “subjective” measures of auditory streaming (i.e., reports of perceived segregation) were collected in the same listeners under corresponding stimulus conditions.

EXPERIMENT 1: JUDGMENTS OF PERCEIVED SEGREGATION

Rationale

The aim of this experiment was to collect subjective measures of auditory streaming using tone sequences that were as similar as possible to those used in the other two experiments, which were designed specifically to obtain objective psychophysical measures, i.e., thresholds. For reasons that will become apparent later (see “the rationale of experiment 2”), these two other experiments required that the temporal positions of the tones be “jittered” randomly, resulting in temporally irregular sequences. Thus, similar temporal jittering was applied to the tones in this first experiment. Moreover, because the tasks in the two main experiments were designed to encourage either segregation or integration, we were interested in how listeners’ percepts were influenced by what van Noorden (1975) referred to as the listener’s “attentional set.” Van Noorden (1975) showed that while auditory streaming is partly automatic, consistent with Bregman’s (1990) description of it as a “primitive” scene-analysis process, listeners nonetheless have some degree of control over the percept. In particular, he showed that when listeners are instructed to try to “hold on” to the percept of a single stream, compulsory segregation occurs at a larger frequency separation than when the listener is actively trying to hear out one of “two streams”. In addition, his results indicate that when the listener is actively trying to segregate, the frequency segregation at which stream segregation is perceived increases markedly as the rate of tone presentation decreases; in contrast, if the listener is actively trying to hear the stimulus as a coherent stream, the frequency separation at which segregation can be perceived is essentially independent of the tone-presentation rate. Therefore, in this experiment, we measured judgments of perceived stream segregation as a function of three stimulus parameters—frequency separation, presentation pace, and number of tones—under three different instruction conditions: neutral instructions, integration-promoting instructions, and segregation-promoting instructions.

Methods

Eleven listeners (seven female and four male, aged 19 to 24 years) took part in this experiment. Prior to inclusion in the study, listeners provided written informed consent, and pure-tone audiometry was performed. All listeners had normal hearing, defined as pure-tone hearing thresholds of less than 20 dB HL at octave frequencies between 250 and 8,000 Hz.

The stimuli were sequences of tone triplets, ABA, where A and B represent pure tones of (usually) different frequencies. The pure tones were gated on and off with 20-ms cosine-squared ramps. Three stimulus parameters were varied: the frequency separation between the A and B tones, ΔF ; the nominal duration of the inter-tone interval, T ; and the number of triplets in the sequence, N . Depending on the condition being tested, ΔF was equal to 0, 1, 3, 6, 9, or 15 semitones. When ΔF was equal to 0 semitones, the A and B tones had the same frequency (1,000 Hz). The frequency of the B tones was kept constant at 1,000 Hz, while that of the A tones was set ΔF semitones below 1,000 Hz, i.e., at approximately 944, 841, 707, 595, or 420 Hz. Once selected, the value of ΔF was kept constant within a sequence. The parameter T controlled the duration of each tone in the sequence (which equaled T , including the on and off ramps), the nominal (or long-term average) duration of the silent interval between consecutive tones within a triplet (which also equaled T), and the nominal duration of the inter-triplet interval (which equaled $2T$). For reasons that will be explained in a subsequent section, the actual durations of the inter-tone and inter-triplet intervals varied pseudo-randomly across, as well as within, the sequences. The duration of the gap between two consecutive triplets in a sequence could take on a value of either 0 or $4T$ ms, and that of the gap between two consecutive tones within a triplet could be either 0 or $2T$ ms. Due to this random variability, the sequences were temporally irregular. With gaps of a nominal duration of 0 ms, successive tones were perceived as discrete events, even when they were the same frequency, because of the gap introduced by the onset and offset ramps. Depending on the condition, T was equal to 50 or 100 ms, yielding two (average) rates: a fast rate (approximately 9 tones/s) and a slow rate (approximately 4 tones/s). The number of triplets, N , could equal 1, 2, 4, or 8, yielding sequences of different lengths. The combination of these different conditions resulted in a total of 48 stimulus conditions.

This experiment involved three phases. In the first phase, the 48 stimulus conditions resulting from the combination of the three factors, N (4 levels), T (2 levels), and, ΔF (6 levels), were presented eight times each to each listener in completely randomized order. After each sequence presentation, listeners were instructed to report whether, at the end of the sequence, they were hearing a single stream or two separate streams. Since subsequent experiments measured the ability to discriminate temporal changes that always involved the last triplet in the sequence, we were specifically interested in whether listeners heard these sequences as one or “two streams” at the end of each sequence. Therefore, the listeners were instructed

specifically to wait until the end of each sequence before forming their judgment, and to base their response on what they were hearing at the end of the sequence. For each condition, the number of trials on which the listener reported having heard the sequence as “two streams” was divided by 8, the number of trials per condition. The result was used as an estimate of the proportion of trials in which the listener experienced a percept of segregation by the end of the sequence.

The second and third phases of the experiment involved the same stimuli, but different instructions. In one of the two phases (the second for half of the participants, the third for the other half), the listeners were instructed to actively “listen for the high-pitch tones” and to try to hear these tones out from the lower-pitch tones by the end of the current sequence. Whenever they felt they had been successful in hearing out the individual streams, they had to press “2,” to indicate that they heard “two streams”. If, despite their efforts, listeners were unable to hear the high tones as a separate stream, they were told to press “1.” In the other phase of the experiment, the listeners were instructed to try to hold on to percept of a single stream and to press 1 whenever they were still able to do so toward the end of the stimulus sequence, and 2 otherwise.

A Madsen Conera™ Diagnostic Audiometer (GN Otometrics, A/S) was used for pure-tone audiometry. During the experiments proper, stimulus presentation and response collection were controlled using the AFC software package (Stefan Ewert, Universität Oldenburg) under Matlab (The MathWorks, Inc.). The stimuli were generated digitally and played out via a soundcard (LynxStudio L22) with 24-bit resolution and a sampling frequency of 32 kHz. They were presented monaurally to the listener via Sennheiser HD 580 headphones.

Results

The proportions of “two streams” responses that were measured based on the listeners’ subjective reports are shown in Figure 1. The top panel corresponds to the “neutral instructions” condition. The proportion of “two streams” responses increased markedly with ΔF [$F(5, 50)=42.83, p<0.0005, \eta^2=0.811$]. It also increased significantly with N [$F(2, 20)=9.93, p=0.001, \eta^2=0.498$], albeit less markedly; this effect was most evident at intermediate ΔF s. The slower sequences (those with the larger T value) tended to produce fewer “two streams” responses than the faster sequences (with the smaller T value) at the 6- and 9-semitone ΔF [$F(1, 10)=4.37, p=0.063, \eta^2=0.304$].

The middle panel shows the proportion of “two streams” responses that were measured in the condition in which listeners were encouraged to try to hear the sequence as a coherent stream, and to use the two-

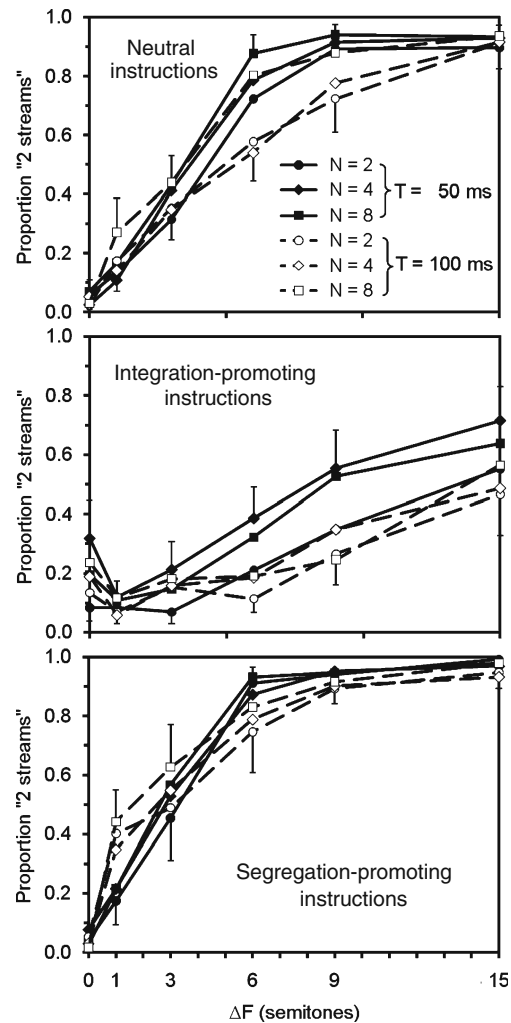


FIG. 1. Mean proportion of “two streams” responses measured as a function of frequency separation (ΔF) in experiment 1. Each panel shows results obtained under a different instructions condition: neutral instructions (*top panel*), integration-promoting instructions (*middle panel*), and segregation-promoting instructions (*lower panel*). Different symbols are used to indicate results obtained using different sequence lengths: $N=2$ (circles), $N=4$ (diamonds), and $N=8$ (squares). Data points corresponding to results obtained using a T of 50 ms (fast presentation rate) are shown as solid symbols connected by *solid lines*. Data points corresponding to results obtained using a T of 100 ms (slow presentation rate) are shown as open symbols connected by *dashed lines*. The *error bars* show standard errors of the mean. To avoid overlap, some *error bars* are not shown.

stream response only when they could not help hearing “two streams”. As expected, fewer “two streams” responses were observed in this condition than in the neutral-instructions condition. Nonetheless, the instructions used in this condition did not eliminate the effect of stimulus parameters. The number of “two streams” responses increased with ΔF [$F(5, 50)=5.79, p=0.018, \eta^2=0.366$], and with N [$F(2, 20)=5.93, p=0.010, \eta^2=0.372$]. Both of these effects were significantly less pronounced at the slower rate than at the faster rate [as indicated by significant interactions between rate and

ΔF , $F(5, 50)=3.29$, $p=0.015$, $\eta^2=0.248$, and between rate and N , $F(5, 50)=3.48$, $p=0.050$, $\eta^2=0.258$].

The bottom panel shows the proportion of two streams responses that were measured in the condition in which listeners were encouraged to try to hear out the high-pitch tones, and to use the “one stream” response only when they could not help but hear the sequence as a coherent stream. The number of “two streams” responses increased with ΔF [$F(5, 50)=60.61$, $p<0.001$, $\eta^2=0.858$], but not with N [$F(2, 20)=1.61$, $p=0.229$, $\eta^2=0.139$]. The effect of ΔF was influenced by the average rate of tone presentation: The slower rate produced less segregation at large ΔF s, but more at small ΔF s [as indicated by a significant interaction between these two factors: $F(5, 50)=5.38$, $p=0.001$, $\eta^2=0.350$].

Discussion

As in previous studies where temporally regular tone sequences have been used (Miller and Heise 1950; van Noorden 1975), the frequency separation between the A and B tones was found to be the main determinant of the listener’s percept. Increases in the proportion of “two streams” responses with increasing ΔF were observed under all three listening conditions. Also consistent with earlier findings (van Noorden 1975), the effect of ΔF was less pronounced in the condition in which listeners were instructed to try to hear the sequence as a single stream, compared with the condition in which listeners were encouraged to hear out the high-pitch tones, and with the neutral-instructions condition.

The finding that the effect of ΔF was similar for the neutral-instructions condition as for the segregation-promoting-instructions condition is surprising. Based on earlier findings (van Noorden 1975), we expected that instructing the listeners to actively try to “hear out” the higher-pitch tones would encourage segregation and result in more “two streams” responses than in the neutral-instructions condition—especially at intermediate ΔF s. The data did not confirm this prediction. A possible explanation for this outcome is that the listeners did not follow our instructions to try to actively hear out the higher-pitch tone. Another possible explanation is that the listeners may have been inclined to hear out the high-pitch (or the low-pitch) tones, even when they were not instructed to do so. This would explain the relatively high proportions of “two streams” responses measured at large frequency separations in the neutral-instructions condition. Consistent with this, after the experiment, some of the listeners reported that the higher-pitch “bird” had grabbed their attention, and that they were primed to attend selectively to it, even before we instructed them to do so.

Based on data in the literature, we also expected higher proportions of “two streams” responses at the faster tone-presentation rate, compared with the lower rate (van Noorden 1975). The data did not confirm this expectation. Interestingly, a smaller effect of presentation rate was observed in the condition in which listeners were encouraged to segregate than in the other two listening conditions. This outcome is compatible with the results of van Noorden (1975), which show a reduced effect of tone presentation rate when listeners are actively trying to segregate.

Finally, based on the results of earlier studies, which showed that stream segregation usually “builds up” over time (van Noorden 1975; Bregman 1978; Anstis and Saida 1985; Carlyon et al. 2001; Cusack et al. 2004; Micheyl et al. 2005a; Pressnitzer et al. 2008), we expected the proportion of “two streams” responses to increase as a function of sequence length. Instead, relatively large proportions of “two streams” responses were observed in response to short stimulus sequences that contained only two triplets, as long as the frequency separation was relatively large. In the “General discussion”, we consider possible explanations for this outcome.

To summarize, subjective measures of streaming depended on ΔF and T in a way that was consistent with our expectations based on earlier data, but they failed to show large effects of N . Moreover, the way in which these measures were influenced by listening instructions was not entirely consistent with our expectations. The proportion of “two streams” responses was not larger when the instructions were to segregate than when instructions were neutral. Our above discussion of these outcomes underscores some of the difficulties in interpreting subjective measures of auditory streaming. One of these difficulties relates to the lack of objective criteria to determine whether listeners followed the instructions that were given to them—or, when they were not given specific instructions, what they were listening for. Another difficulty stems from the fact that, just as yes/no responses, one stream/two streams responses are highly susceptible to individual biases (Green and Swets 1966; Macmillan and Creelman 2005), which can be either sensory (e.g., an a priori inclination to try to segregate) and/or decisional (e.g., a lower criterion for responding “two streams”). These caveats should be kept in mind when relating subjective and objective psychophysical measures of auditory streaming.

EXPERIMENT 2: OBJECTIVE MEASURE OF STREAM INTEGRATION

Rationale

The aim of the second experiment was to measure thresholds in a temporal-discrimination task in which

performance was hypothesized to depend on listeners' perception of the stimulus sequence as a coherent stream. The task was inspired by earlier studies involving auditory streaming and the perception of temporal relationships between sounds (van Noorden 1975; Vliegen et al. 1999; Roberts et al. 2002, 2008; Micheyl and Oxenham 2010). The general principle of these studies involves a temporal shifting of the B tones relative to the A tones in an AB or ABA stimulus sequence. Whereas in previous studies the shift was either applied to all B tones in the sequence or introduced progressively, here it was applied only to the last B tone in the sequence. In all other triplets within the sequence, the B tone was temporally centered between the two adjacent A tones. In order to encourage listeners to rely on within-triplet comparisons of timing between the A and B tones, rather than on across-triplet comparisons of the inter-B-tone timing, the inter-triplet gap was randomly jittered.

Method

Seven of the 11 listeners who had taken part in experiment 1 were retained to take part in this and the following experiments. The selection of these seven listeners did not involve any particular criteria, other than their schedules permitting participation in the experiment two to three times a week, 2 h each time, within the timeframe of the study.

The stimuli were similar to those used in experiment 1. The only differences were as follows. The timing of the B tone relative to the two A tones in each triplet was constant, up to the last triplet, where the B tone was shifted by $\pm\Delta t$, relative to the center-point between the two adjacent A tones in the same triplet, as illustrated in Figure 2A. The only randomly varying quantity in this experiment was the duration of the silent interval between successive triplets. This was produced by shifting the onset time of each triplet randomly by $\pm T$ ms (with equal probability), independently for each triplet. As a result, the time interval between any two triplets in the sequence could equal 0 ms with probability 0.5, $2T$ ms with probability 0.25, or $4T$ ms with probability 0.25. Note that as in experiment 1, a gap of 0 ms at the zero-amplitude points was in fact a gap at 25.4 ms at the half-amplitude points because of the onset and offset ramps applied to each tone. As in the previous experiment, T was set to either 50 or 100 ms, yielding a fast-rate and a slow-rate condition.

In this experiment, the task of the listener was to focus on the last triplet of a stimulus sequence and to indicate whether the B tone in that triplet was shifted forward or backward in time relative to the two adjacent A tones within the same triplet. In practice, to make the instructions easy to understand, listeners

were asked to report whether the temporal pattern (rhythm) evoked by the last triplet heard was more similar to "AB-A" or "A-BA," where the dash denotes a silent gap. The last triplet was the only triplet in the sequence in which the duration of the silent intervals before and after the B tone were not equal. The temporal jitter in the time interval between successive triplets made the timing between successive B tones irregular, thereby rendering the time between the last and the penultimate B tone unreliable as a cue. Therefore, in this task, it was advantageous to compare the timing of the A and B tones in the last triplet. Listeners were not explicitly told in advance how many triplets the sequence would contain. However, this number was constant within each "run" of the adaptive threshold-tracking procedure.

The smallest temporal shift of the B tone for which listeners could correctly discriminate the shift direction on 70.7% of the trials was measured using a transformed two-down, one-up adaptive procedure (Levitt 1971). The tracking variable, Δt , was expressed as a percentage of T . At the beginning of a "run" of the adaptive procedure, Δt was set to 100% of T . It was increased or decreased by 10 percentage points until the first reversal (going from "up" to "down") in the direction of the adaptive staircase, then adjusted by five percentage points over the next two reversals, and finally by 2.5 percentage points over the last six reversals, at which point the procedure was terminated. Threshold was estimated as the mean of the Δt values at the last six reversals. The direction of the temporal shift (forward or backward) applied to the last B tone on the current trial was selected at random, with the two possible directions having equal a priori probabilities. Listeners gave their responses by pressing 1 or 2 on a computer keyboard, with 1 corresponding to a forward shift, and 2 to a backward shift. Following each response, feedback was provided in the form of a message ("correct" or "wrong") on the computer screen. Between three and six threshold measurements per condition were obtained, depending on the listener's availability. For each listener and condition, the three "best" (i.e., lowest) thresholds were averaged, and the result was used as the final threshold estimate (see Appendix).

Results and discussion

The results of this experiment are shown in Figure 3. Thresholds increased with both ΔF [$F(5, 30)=21.47$, $p<0.001$, $\eta^2=0.782$] and N [$F(3, 18)=23.38$, $p<0.001$, $\eta^2=0.796$]. They were significantly larger on average at

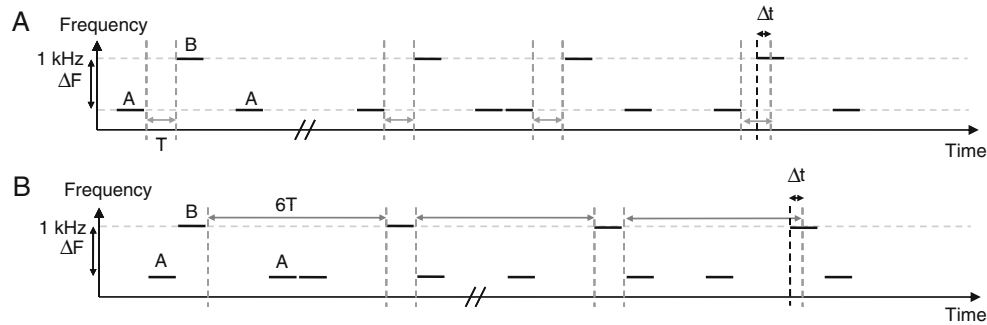


FIG. 2. Schematic spectrograms of example stimulus sequences presented on a trial in experiments 2 and 3. The A and B tones are labeled only for the first triplet. Two of the three main stimulus parameters, frequency separation (ΔF) and the within-triplet inter-tone interval (T), are indicated explicitly on the schema. The third parameter, number of triplets (N), is not shown; in these examples, it was equal to 4. **A** In experiment 2, the duration of the interval between the A and B tones within each triplet, which is labeled T and shown using *double gray arrows*, was constant throughout the sequence, except in the last triplet, where it was either reduced or increased (at random with equal probability) by Δt . In this example,

the interval was *decreased*, so that the B tone was shifted toward the leading A tone and away from the trailing A tone. On other trials, the shift could be in the opposite direction. In this experiment, the inter-triplet interval was varied independently for each pair of triplets, including the last pair. **B** In experiment 3, the duration of the interval between consecutive B tones, which is labeled $6T$ and shown using *double gray arrows*, was constant throughout the sequence except for the last pair of B tones, where it was reduced or increased (at random with equal probability) by Δt . In this experiment, the timing of the A tones relative to the B tones was varied, with both A tones from a triplet shifted forward or backward coherently by T ms.

the faster rate than at the slower rate [$F(1, 6)=6.33$, $p=0.045$, $\eta^2=0.514$].

This pattern of results is consistent with the conclusions of earlier studies (Vliegen et al. 1999; Roberts et al. 2002, 2008), that compulsory stream segregation hampers listeners' ability to correctly identify the direction of temporal shifts across frequency. The increases in thresholds that were observed in this experiment as a function of ΔF and T can be explained, at least qualitatively, based on the effect of these parameters on stream segregation, which were seen directly in experiment 1.

The increase in thresholds with increasing N observed in the current experiment is consistent with our original hypothesis that in this task, an increase in compulsory segregation with increasing sequence length would limit performance. However, no clear effect of N was found in the "integration" condition of experiment 1, suggesting that with these short stimulus sequences, N had no consistent effect on stream segregation. Moreover, at the fast rate, thresholds increased as N was increased from one to two, even in the zero ΔF condition. This effect cannot be due to stream segregation because listeners presumably did not experience segregation when there was no ΔF . A tentative explanation is that at the fast rate, the introduction of "precursor" triplets before the target triplet had a distracting influence. The fact that this effect was only present at the fast rate could be due to the precursor and target triplets being closer in time—and thus more likely to interfere with each other in perception—at the fast rate than at the slow rate. Another observation that suggests that the effects observed in this experiment cannot be

explained entirely in terms of auditory streaming is that for $N>1$, thresholds were usually higher at the faster rate than at the slower rate, even at zero ΔF 's. This could be due to the longer tone durations and longer inter-tone intervals, which may have allowed listeners to better focus selectively on the last triplet. Thus, some but not all of the effects can be explained by auditory streaming. We return to this important point in the "General discussion".

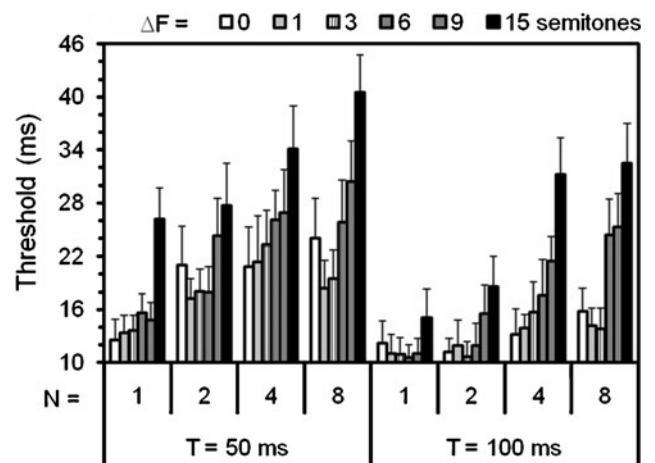


FIG. 3. Mean thresholds in experiment 2. The *different bar shadings* correspond to different ΔF 's, as indicated in the *top legend*. The *numbers 1, 2, 4, and 8* along the *abscissa* refer to the number of tones in the sequence, N . *Bars on the left half of the plot* indicate thresholds obtained with a T of 50 ms (fast presentation rate). *Bars on the right half of the plot* indicate thresholds obtained with a T of 100 ms (slow presentation rate). The *error bars* indicate plus one standard error above the mean.

EXPERIMENT 3: OBJECTIVE MEASURE OF STREAM SEGREGATION

Rationale

Whereas the task used in experiment 2 was designed in such a way that stream segregation would impede performance, the current experiment was designed so that segregation should improve performance. Listeners were asked to discriminate between a forward and a backward shift of the last B tone relative to the preceding B tone(s) in the same sequence. At the same time, the timing of the A tones was jittered, so that listeners would derive little or no useful information from timing comparisons between the A and B tones. Thus, whereas in experiment 2 listeners were asked to compare the timing of the A and B tones, here, they were instructed to ignore the A tones and to focus on judging the relative timing of the B tones. We hypothesized that listeners would find it easier to do so in conditions that facilitated the perceptual segregation of the A and B tones, because in such conditions it would be possible to listen selectively to the stream of B tones. Accordingly, we predicted that thresholds would improve (i.e., become smaller) as the frequency separation and the sequence length increased, and that thresholds would be smaller in the faster condition than in the slower condition. In other words, we expected a pattern of results opposite to that observed in experiment 2.

Methods

The same seven listeners who took part in experiment 2 also took part in the current experiment. The two experiments were run in randomized order, such that four of the listeners completed experiment 2 before they performed experiment 3, while the other three listeners did the opposite.

The stimuli for this experiment are illustrated schematically in Figure 2B. They were similar to those used in the previous two experiments, with the following exceptions. First, only sequence lengths (N) of 2, 4, and 8 were used here. Second, except for the last (or the only) two B tones in the sequence, the time interval between consecutive B tones was kept constant; this interval (measured from the offset of one B tone to the onset of the next) was fixed at $6T$ where, as before, T was set to either 50 or 100 ms in order to produce two presentation-rate conditions. Third, the interval from the offset of the penultimate B tone to the onset of the last B tone was varied adaptively, according to the same two-down one-up tracking rule used in experiment 2; it was equal to $6T \pm \Delta t$, where Δt was varied adaptively during the course of a block of trials, based on the listener's responses. Fourth, the timing of the A tones was

shifted randomly forward or backward by T . The two A tones within each triplet were shifted in the same direction and by the same amount. Finally, the " $\Delta F=0$ " condition from experiment 2 was replaced by a "no A tones" condition. As the name indicates, in this condition, the amplitude of the A tones was set to zero. The temporal characteristics of the B tones were the same as in the other conditions. This condition was run to assess the influence of the A tones on performance in this task. We reasoned that thresholds would be lowest in this condition, and that this would provide a baseline against which the thresholds measured in other conditions (where A tones were present) could be compared.

The task for listeners was to indicate after each trial whether the last B tone in the sequence was "early" or "late" or, equivalently, whether the time interval between the last two B tones in the sequence was "shorter" or "longer." Listeners gave their responses by pressing 1 or 2 on a computer keyboard, with 1 corresponding to early (or shorter), and 2 corresponding to late (or longer). Following each response, feedback was provided in the form of a message ("correct" or "wrong") on the computer screen. Between three and six threshold measurements per condition were obtained, depending on the listener's availability. For each listener and condition, the three "best" (i.e., lowest) thresholds were averaged, as in experiment 2, and the result was used as the final threshold estimate.

Results and discussion

The results of this experiment are shown in Figure 4. As predicted under the hypothesis that stream segregation should help task performance, thresholds decreased as ΔF increased [$F(4, 24)=8.94$, $p<0.002$, $\eta^2=0.598$], and they were higher on average for the slow sequences than for the fast sequences [$F(1, 6)=44.45$, $p=0.001$, $\eta^2=0.881$]. However, the effect of rate cannot be due entirely to stream segregation because it was present even in the control condition with no A tones [$F(1, 6)=13.29$, $p=0.011$, $\eta^2=0.689$], a condition in which the stimuli could only be perceived as a single stream. Larger thresholds in slow-rate conditions than in fast-rate conditions would be expected if the accuracy of interval-duration discrimination decreased as the duration of the baseline interval increases, as expected from Weber's law. For $N=2$, the mean threshold measured in the slow-rate condition with no A tones was 1.82 times larger than the mean threshold measured in the corresponding fast-rate condition. For longer sequences, the ratio decreased to 1.35 (for $N=4$) and 1.20 (for $N=8$). To determine whether these changes in thresholds in the absence of

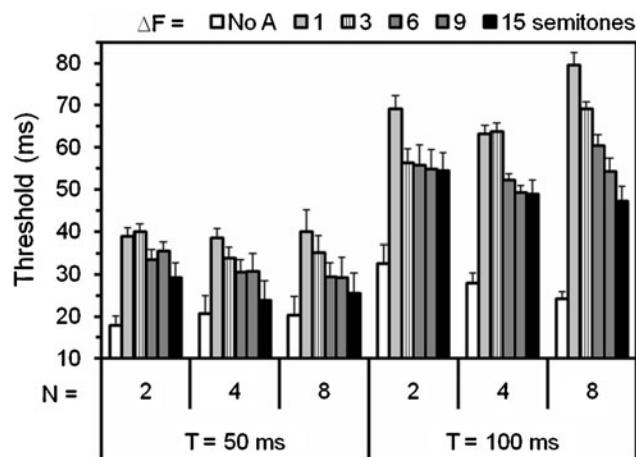


FIG. 4. Mean thresholds in experiment 3. The different bar shadings correspond to different ΔF conditions, as indicated in the top legend. Open bars indicate thresholds measured in the absence of A tones. The numbers 2, 4, and 8 along the abscissa refer to the number of elements (ABA tone triplets or B-B tone pairs) in the sequence, N . Bars in the left half of the plot indicate thresholds obtained with a T of 50 ms (fast presentation rate). Bars in the right half of the plot indicate thresholds obtained with a T of 100 ms (slow presentation rate). The error bars indicate plus one standard error above the mean.

the A tones could account for the differences in thresholds between the slow- and fast-rate conditions in the presence of the A tones, we divided the thresholds measured in the slow-rate conditions by the corresponding ratio. Even after this correction, thresholds were still significantly larger in the slow-rate condition than in the fast-rate condition [$F(1, 6) = 6.11$, $p = 0.048$, $\eta^2 = 0.505$]. Therefore, the difference in thresholds between the slow- and fast-rate conditions in the presence of the A tones cannot be explained simply by the observed difference in thresholds observed in the absence of the A tones.

Based on the hypothesis that stream segregation would facilitate task performance in this experiment, we had initially hypothesized that thresholds would improve as the sequence length, N , increased. The data did not support this hypothesis [$F(2, 12) = 1.03$, $p = 0.380$, $\eta^2 = 0.146$]. While this lack of effect of sequence length was not expected a priori, it is consistent with the subjective data from experiment 1, which also showed no significant effect of N in conditions in which listeners were encouraged to segregate and follow the high-pitch stream. This suggests that when listeners were actively trying to segregate sounds in the current study, stream segregation occurred almost instantaneously after the start of the stimulus sequence. Under these conditions, the segregation “build-up” effect, which has been observed in other studies under neutral listening conditions (van Noorden 1975; Bregman 1978; Anstis and Saida 1985; Carlyon et al. 2001; Cusack et al. 2004; Micheyl et al. 2005a; Pressnitzer et al. 2008) and conditions that encouraged integration (Roberts et al. 2008), appears

to have. However, the lack of a significant effect of N in this experiment should not be interpreted as evidence that build-up effects cannot be obtained under conditions that promote stream segregation. In fact, in an earlier study, an effect of sequence length on thresholds was observed in a within-stream frequency-discrimination task that facilitated stream segregation (Micheyl et al. 2005b). However, as suggested by the authors of that earlier study, while this effect is consistent with the build-up of stream segregation, it may also have been due to other factors. Specifically, the task used in that study required listeners to discriminate a change in the frequency of the final B tone in a repeating ABA sequence. An increase in the number of “precursor” triplets could benefit performance by providing multiple occasions for listeners to “sample” the frequency of the B tone before that frequency was shifted (up or down) in the last triplet. This “multiple looks” mechanism (Swets et al. 1959) might have been expected to play a role in the current experiment also, in that more triplets resulted in more opportunities to sample the “standard” time interval between successive B tones. However, as the results showed no effect of N , it does not seem that the multiple-looks mechanism played a substantial role in the current experiment.

GENERAL DISCUSSION

Relationship between subjective reports and performance-based measures

One of the motivations of this study was to examine relationships between judgments of perceived segregation and temporal-discrimination thresholds measured using similar stimuli, in the same listeners. The results described above already revealed some consistent trends in the dependence of these two types of measures (averaged across all listeners) as a function of stimulus parameters. However, ideally, experimenters would like to be able to use one type of measure to infer the other measure, for a particular stimulus condition, in a given listener. Therefore, it is important to examine in more detail how the two types of measures are statistically related to each other at the individual level.

To this aim, we computed non-parametric correlation coefficients (Spearman’s ρ) between the proportions of “two streams” responses measured in experiment 1 and the thresholds measured in experiments 2 and 3. This was done separately for each of the seven listeners who took part in all three experiments, as well as for the mean data across these seven listeners. In addition, the correlations were computed for both matching and non-matching conditions between the experiments. For example, thresholds

measured in experiment 2, which used a task in which stream integration was expected to facilitate performance, were first correlated with the proportions of “two streams” responses measured in the integration-promoting instructions condition of experiment 1 (matching conditions). Then, the same thresholds were correlated with the proportions of “two streams” responses measured in the segregation-promoting instructions of experiment 1 (non-matching conditions).

The results of these correlation analyses are listed in Tables 1 and 2. For experiment 2 (Table 1), significant correlations with the proportions of “two streams” responses measured under matching instructions in experiment 1 were observed in four out of the seven listeners, as well as for the data averaged across all listeners. When the correlations were computed using the proportions of “two streams” responses measured under non-matching instructions, statistically significant correlations were observed in only one out of seven listeners. However, the correlation computed based on the mean data was significant, and that correlation was not significantly smaller than the correlation obtained using matching instructions. Consistent with the trends described above (Fig. 3), these correlations were all positive. For experiment 3 (Table 2), significant correlations were observed for only two listeners, and only when non-matching instructions were used. However, when the data were averaged across listeners, significant correlations were observed for both matching and non-matching

TABLE 1

Correlations between the proportion of “two streams” responses measured in experiment 1 and the discrimination thresholds measured in experiment 2

	<i>Matching instructions</i>	<i>Non-matching instructions</i>
Listener 1	$\rho = -0.15, p = 0.199$	$\rho = 0.21, p = 0.121$
Listener 2	$\rho = 0.49, p = 0.002^a$	$\rho = 0.44, p = 0.006^a$
Listener 3	$\rho = 0.59, p < 0.001^a$	$\rho = 0.27, p = 0.068$
Listener 4	$\rho = 0.34, p = 0.027$	$\rho = 0.18, p = 0.168$
Listener 5	$\rho = 0.68, p = 0.191$	$\rho = 0.25, p = 0.080$
Listener 6	$\rho = 0.16, p = 0.001^a$	$\rho = 0.20, p = 0.138$
Listener 7	$\rho = 0.50, p < 0.001^a$	$\rho = 0.36, p = 0.020$
Mean	$\rho = 0.69, p < 0.001^a$	$\rho = 0.61, p < 0.001^a$

The ρ values list non-parametric (Spearman) correlation coefficients. The one-sided p values were computed as $1 - \Phi(\tanh^{-1}(|\rho|)\sqrt{n-3}/1.06)$, where Φ denotes the cumulative standard normal function, \tanh^{-1} denotes the hyperbolic arc-tangent function, and n was equal to 30 (Fieller 1957). The “Matching instructions” column lists correlations with the proportions of two streams responses measured under integration-promoting instructions in experiment 1. The “Non-matching instructions” column lists correlations with the proportions of “two streams” responses measured under segregation-promoting instructions in experiment 1. The values in the last row were computed based on the mean proportions of “two streams” responses and mean discrimination thresholds, averaged across the seven listeners

^aAll p values lower than 0.05/16=0.0031, indicating statistically significant correlations after the application of Bonferroni’s correction

TABLE 2

Correlations between the proportion of “two streams” responses measured in experiment 1 and the discrimination thresholds measured in experiment 3

	<i>Matching instructions</i>	<i>Non-matching instructions</i>
Listener 1	$\rho = -0.13, p = 0.256$	$\rho = 0.10, p = 0.317$
Listener 2	$\rho = -0.43, p = 0.012$	$\rho = -0.53, p = 0.002^a$
Listener 3	$\rho = -0.28, p = 0.076$	$\rho = -0.09, p = 0.336$
Listener 4	$\rho = -0.08, p = 0.341$	$\rho = -0.30, p = 0.066$
Listener 5	$\rho = -0.16, p = 0.212$	$\rho = -0.73, p < 0.001^a$
Listener 6	–	$\rho = 0.35, p = 0.038$
Listener 7	$\rho = -0.29, p = 0.070$	$\rho = -0.44, p = 0.011$
Mean	$\rho = -0.46, p < 0.008^a$	$\rho = -0.59, p < 0.001^a$

The “Matching instructions” column lists correlations with the proportions of “two streams” responses measured under segregation-promoting instructions in experiment 1. The “Non-matching instructions” column lists correlations with the proportions of “two streams” responses measured under integration-promoting instructions in experiment 1. The missing values for listener 6 in the matching condition are due to the fact that when instructed to actively try to hear out the higher-pitch tones in experiment 1, this listener always responded “two streams” for all but the zero frequency separations. Since zero frequency separations were not tested in experiment 3, proportions of two streams responses corresponding to zero frequency separations could not be included in the correlations shown here. Otherwise, as Table 1

^aAll p values lower than 0.05/16=0.0031, indicating statistically significant correlations after the application of Bonferroni’s correction

instructions, and the correlations did not differ significantly from each other. Consistent with the trends described above (Fig. 4), these correlations were all negative.

In summary, significant correlations were observed between the mean proportions of “two streams” responses and the mean discrimination thresholds averaged across the listeners who took part in all three experiments. These “global” correlations did not depend significantly on whether matching or non-matching instructions were considered. Significant correlations were observed for some individual listeners, especially for experiment 2 and when matching instructions were considered. However, these correlations were not observed with sufficient consistency to allow experimenters to use measured proportions of “two streams” responses to predict discrimination thresholds, or vice versa, in individual listeners, using the current experimental procedures.

The failure to find robust correlations between temporal-discrimination thresholds and subjective measures of auditory streaming at the individual level is perhaps not surprising, considering that the former were influenced by factors beyond the perceptual organization of the stimulus sequence. The listener’s ability to discriminate relatively small changes in the duration of temporal intervals was probably one such factor. For instance, in experiment 3, thresholds were larger on average in the slow-rate than in the fast-rate conditions, even when the A tones were absent and the stimulus sequences could only be perceived as a

single stream. This effect could be due to an increase in the internal noise associated with the sensory representation of temporal intervals as interval duration increases (Creelman 1964; Allan and Kristofferson 1974). Another factor that may have played a role in determining listeners' performance in experiments 2 and 3 relates to listeners' limited ability to listen selectively in time (Wright and Dai 1994), which may be thought of as a detrimental by-product of temporal integration (Micheyl and Carlyon 1998). For instance, the finding of larger thresholds for $N=2$ than for $N=1$ in the fast-rate condition of experiment 2 may have been due to listeners not being able to completely ignore the irrelevant tones that preceded the target triplet, when these tones were not sufficiently well separated in time from each other.

Separating the contributions of factors related to auditory streaming and of other factors is an important goal for future studies. This may require—or be facilitated by—the formulation of mathematical models of the effects of these factors on performance in the tasks that are used to obtain objective measures of streaming. For the temporal-perception tasks considered here, this would entail combining models of duration discrimination (e.g., Creelman 1964; Allan and Kristofferson 1974) with models of sequential organization (e.g., Beauvois and Meddis 1996; McCabe and Denham 1997; Kanwal et al. 2003; Micheyl et al. 2005a). Awaiting such studies, the conclusion at present is that objective psychophysical measures of streaming obtained using the temporal-discrimination tasks described in this study cannot be used to infer the probability that a given listener perceives a particular stimulus sequence as one stream or two. However, the significant correlations that were observed between the mean discrimination thresholds and the proportions of “two streams” responses suggest that the stimuli and tasks described here could be useful in future behavioral or neurophysiological studies that seek to investigate neural correlates of auditory streaming. We come back to this last point in the final section of the discussion.

Effects of stimulus and task on the build-up of auditory stream segregation

One puzzling aspect of the results of the current study relates to the lack of systematic “build-up” effects in both the judgments of perceived segregation (experiment 1) and the thresholds (experiments 2 and 3). Although significant effects of sequence length were occasionally observed in experiments 1 and 2, these effects were often small and unsystematic. Below, we consider various tentative explanations for this outcome.

A first possible explanation is that in most previous studies showing a build-up effect the tones were temporally regular, whereas in the current study the tones were temporally jittered. Such jittering may have altered the build-up. Okada and Kashino (2008) found that the rate of perceptual alternations in auditory streaming was reduced by the application of temporal jittering. On the other hand, two earlier studies found no significant effect of temporal regularity and sequence predictability on auditory streaming (French-St George and Bregman 1989; Rogers and Bregman 1993). In particular, French-St George and Bregman (1989) found no significant effect of temporal jittering on the mean duration over which listeners could hold on to a single-stream percept. More recently, Roberts et al. (2008) observed a build-up effect in a temporal-discrimination task in which the timing of the B tones changed progressively over the course of a long tone sequence (Roberts et al. 2008). Therefore, the use of temporally irregular sequences does not appear to be a likely explanation for the lack of clear and systematic build-up effect in the current experiment.

A second possible explanation relates to our use of relatively short stimulus sequences: a few seconds or less, compared with 10 s or more in previous studies (e.g., Anstis and Saida 1985; Carlyon et al. 2001; Micheyl et al. 2005a, b). This could affect the build-up in various ways. For instance, it may have encouraged the listeners in this study to pay close attention at the onset of each sequence, and to form perceptual judgments relatively rapidly—since the listeners knew that they would not have ample opportunity to listen to the stimulus before they had to report their judgment. This enhanced “preparedness” may have considerably accelerated or by-passed the build-up, such that segregation was perceived after the first or second triplet. Consistent with this, Pressnitzer (2008) observed faster build-up for short (10-s) sequences than for longer (60-s) sequences when the sequence duration was known in advance to the listener—indicating that the time course of the build-up is influenced by listeners' expectations regarding stimulus duration. It is not currently known whether this effect is mediated by decisional factors (i.e., a change in the listener's decision criterion depending on expected stimulus duration), or by sensory factors under the control of the listener's attention or expectations (e.g., a modulation of the time course of neural adaptation depending on attention or expected stimulus duration). The spontaneous adoption of a low criterion for responding “two streams,” even in the condition where the listeners were not instructed to actively try to separate streams, would be consistent with the finding of similarly high proportions of two-stream responses in this condition and

the condition where listeners were instructed to actively try to separate streams, in experiment 1.

Finally, a third factor that may have contributed to weak or absent build-up effects in this study is that whereas in most studies showing build-up listeners provided their responses *during the course* of an ongoing stimulus sequence, in the current experiment, the listeners had to indicate their percept *after* the end of the stimulus sequence. This gave listeners the opportunity to “rehearse” the just-heard stimulus before they chose their response. Mental rehearsal of short stimuli, and the additional time to decide, may have influenced the decision outcome. Additional study is required in order to clarify how the build-up of stream segregation is influenced by these various factors.

Possible applications to behavioral and neurophysiological studies of auditory streaming

One of the long-term goals of the research program in the context of which this study was performed is to develop tasks that can be used to study the perceptual organization of sound sequences in non-human species, and compare results from them to measures obtained using similar stimuli and tasks in humans. While streaming percepts can be determined directly and relatively simply in human listeners, who can report what they perceive, the study of auditory streaming in other species depends critically on the development of behavioral tasks in which performance depends on, or at least covaries with, the perceived organization of the stimulus. There have been a few behavioral studies of auditory streaming in non-human species, including birds (Hulse et al. 1997; MacDougall-Shackleton et al. 1998), fish (Fay 1998, 2000), monkeys (Izumi 2002), and ferrets (Ma et al. 2010); recent reviews of these findings can be found in Bee and Micheyl (2008) and Fay (2008). In most of these studies, the behavioral tasks used to measure auditory streaming either encouraged segregation, or favored neither segregation nor integration, corresponding to the neutral-instructions condition of our experiment 1. The two complementary tasks described here may be helpful in expanding the array of behavioral techniques available for investigating auditory stream segregation and integration in animals. Results in the animal-behavior literature suggest that the temporal abilities that are required to perform these tasks are present in various non-human species. In particular, gap-duration discrimination and gap-detection thresholds have already been measured in rats (Church et al. 1976), birds (Hienz et al. 1980; Maier and Klump 1990), and monkeys (Sinnott et al. 1987), and the results

indicate that thresholds in these species are often as good, if not better, than those measured in humans.

Another area in which the present results might be useful relates to the search for neural correlates of auditory streaming. In the past decade, several studies have investigated single- or multi-unit neural correlates of auditory streaming in mammals (Fishman et al. 2001; Kanwal et al. 2003; Fishman et al. 2004; Micheyl et al. 2005a; Pressnitzer et al. 2008; Elhilali et al. 2009), birds (Bee and Klump 2004, 2005; Itatani and Klump 2009; Bee et al. 2010), and even insects (Schul and Sheridan 2006). While these studies have identified putative neural mechanisms of auditory streaming, their conclusions have so far been limited by the lack of behavioral measures concomitant to the neural recordings. To the extent that the two temporal-perception tasks described here can be applied in non-human animals, they could prove useful in identifying neural correlates of auditory streaming at the single- or multi-unit level. Specifically, by comparing neural activation patterns during the performance of one of the two types of tasks with neural activation patterns recorded during the performance of the other type of task, researchers might be able to differentiate neural responses that are associated with perceived grouping from neural responses that reflect perceived segregation, in a way that minimizes stimulus-related differences (see Logothetis and Schall 1989; Parker and Newsome 1998).

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health, National Institute on Deafness and other Communication Disorders Grant R01 DC 07657. Nathan Heggem, then a summer student of the REU Program in the Behavioral and Cognitive Sciences organized by Celia Wolk Gershenson and Herb Pick, Department of Psychology, University of Minnesota, helped with data collection. Barbara Shinn-Cunningham and two anonymous reviewers provided many helpful comments on an earlier version of the manuscript.

APPENDIX

The decision to retain only the three best thresholds was motivated by the following considerations. First, we noticed that some of the measured thresholds were unusually large compared with those obtained by the same listener on other tracks, in the same stimulus condition. Based on discussions with our listeners, these unusually large thresholds appeared to correspond to dips in the listener’s concentration. Even though the listeners were encouraged to take

breaks whenever they felt that they needed them, occasional attentional lapses are difficult to avoid in repetitive and protracted psychophysical experiments such as those described here. Although this problem is often ignored, or considered negligible, previous studies indicate that not taking effects of inattention into account can result in threshold overestimation (e.g., Wightman et al. 1989; Wightman and Allen 1994; Green 1995). Wightman et al. (1989) found that in the presence of inattention, the true underlying threshold was actually closer to the best measured threshold than to the mean of the measured thresholds. This finding prompted us to perform Monte Carlo simulations to examine the influence of inattention on the theoretical distribution of threshold estimates produced by the two-down one-up adaptive tracking rule used here. Following previous investigators (Green 1995; Wightman et al. 1989; Wightman and Allen 1994; Viemeister and Schlauch 1994), we modeled attentional lapses as a Bernoulli process, with guesses (a 0.5 probability of correct response) occurring on a randomly selected subset of trials. The proportion of “guessing” trials was controlled by the lapse-rate probability, λ . The larger this probability, the smaller the asymptotic probability of a correct response in the assumed psychometric function,

$$P(\Delta t) = \frac{\lambda}{2} + (1 - \lambda)\Phi\left(\frac{\Phi^{-1}(0.707)\Delta t}{\theta}\right) \quad (1)$$

where Δt denotes the difference in gap duration between the standard and comparison intervals, expressed as a percentage of the base duration T , Φ and Φ^{-1} denote the cumulative normal function and its inverse, and θ denotes the assumed true threshold, i.e., the value of Δt yielding 70.7% of correct responses when $\lambda=0$ (Green 1995). Empirical studies of psychometric functions typically show lapse rates between 0 and 0.2 in young adult listeners. (Note that a lapse rate of 0.2 produces a psychometric function with an upper asymptote at 90% correct because, on average, a correct response still occurs on half of the trials on which the participant guesses.) Accordingly, in the simulations, we used λ values of 0.1 and 0.2. Simulations were performed for different values of θ , ranging from 10% to 90% of T , in steps of 10 percentage points. For each setting of λ and θ , 1,000 simulated tracks were produced, yielding 1,000 threshold estimates. From these 1,000 threshold estimates, 10,000 samples of six values were drawn at random (with replacement), simulating the measurement of six thresholds per condition in the psychoacoustical experiment. For each sample, two statistics were computed: the mean of all six values in the sample and the mean of the three smallest thresholds

in the sample. The latter statistic (mean of the three best) was found to be closer to the true threshold more often than the former (simple mean). For instance, with a lapse rate of 0.1 and a true threshold of 30% of T , the simple mean was equal to 38% while the mean of the best three was equal to 28%. With a lapse rate of 0.2, the simple mean was equal to 45% while the mean of the best three was 35%. As the value of the true threshold increased, the simple mean came closer to the true threshold than the mean of the best three thresholds, i.e., for large thresholds, the bias in the simple mean became smaller than the bias in the mean of the three best thresholds. However, on average across all tested values of the true threshold (from 10% to 90%) and lapse rates (0.1 and 0.2), the mean of the best three thresholds was closer to the true threshold than the simple mean. Another way of stating this is that while the simple mean of all measured thresholds usually over-estimated the true threshold, the mean of the best thresholds usually showed a bias in the opposite direction; however, on average, the bias in the latter was less than the bias in the former. Therefore, we decided to use the mean of the three best thresholds rather than the simple mean.

One consequence of our decision to retain only the best three thresholds is that the mean thresholds reported here are slightly smaller and slightly less variable, on average, than those that would have been obtained by averaging all available measurements. We acknowledge that the mean of the three best thresholds is not an unbiased minimum-variance estimator. Deriving an unbiased minimum-variance estimator of the true thresholds in the presence of inattention effects requires precise knowledge of how much inattention was present, how exactly this inattention affected the responses of the participant, and of what the true threshold was. This information is not usually available to the experimenter. Importantly, however, the mean thresholds reported in the current article fall within the range of auditory duration-discrimination thresholds measured in previous studies (Chistovitch 1959; Abel 1972; Divenyi and Danner 1977; Hirsh et al. 1990; Rammsayer and Lima 1991; Drake and Botte 1993; Fitzgibbons and Gordon-Salant 1994; Phillips et al. 1994; Rammsayer 1994; Fitzgibbons and Gordon-Salant 1995; Wright et al. 1997; Vliegen et al. 1999; Roberts et al. 2002; Fitzgibbons et al. 2007).

REFERENCES

- ABEL SM (1972) Discrimination of temporal gaps. *J Acoust Soc Am* 52:519–524
- ALLAN LG, KRISTOFFERSON AB (1974) Psychophysical theories of duration discrimination. *Percept Psychophys* 16:26–34

- ANSTIS S, SAIDA S (1985) Adaptation to auditory streaming of frequency-modulated tones. *J Exp Psychol: Hum Percept Perform* 11:257–271
- BEAUVOIS MW, MEDDIS R (1996) Computer simulation of auditory stream segregation in alternating-tone sequences. *J Acoust Soc Am* 99:2270–2280
- BEE MA, KLUMP GM (2004) Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. *J Neurophysiol* 92:1088–1104
- BEE MA, KLUMP GM (2005) Auditory stream segregation in the songbird forebrain: effects of time intervals on responses to interleaved tone sequences. *Brain Behav Evol* 66:197–214
- BEE MA, MICHEYL C (2008) The cocktail party problem: what is it? How can it be solved? And why should animal behaviorists study it? *J Comp Psychol* 122:235–251
- BEE MA, MICHEYL C, OXENHAM AJ, KLUMP G (2010) Neural adaptation to tone sequences in the songbird forebrain: patterns, determinants, and relation to the build-up of auditory streaming. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol*. doi: 10.1007/s00359-010-0542-4
- BEST V, GALLUN FJ, CARLILE S, SHINN-CUNNINGHAM BG (2008) Binaural interference and auditory grouping. *J Acoust Soc Am* 121:1070–1076
- BEY C, McADAMS S (2002) Schema-based processing in auditory scene analysis. *Percept Psychophys* 64:844–854
- BEY C, McADAMS S (2003) Postrecognition of interleaved melodies as an indirect measure of auditory stream formation. *J Exp Psychol: Hum Percept Perform* 29:267–279
- BREGMAN AS (1978) Auditory streaming is cumulative. *J Exp Psychol: Hum Percept Perform* 4:380–387
- BREGMAN AS (1990) Auditory scene analysis. MIT Press, Cambridge
- BREGMAN AS, AHAD PA, CRUM PA, O'REILLY J (2000) Effects of time intervals and tone durations on auditory stream segregation. *Percept Psychophys* 62:626–636
- BREGMAN AS, AHAD PA, VAN LOON C (2001) Stream segregation of narrow-band noise bursts. *Percept Psychophys* 63:790–797
- CARLYON RP (2004) How the brain separates sounds. *Trends Cogn Sci* 8:465–471
- CARLYON RP, GOCKEL H (2008) Effects of harmonicity and regularity on the perception of sound sources. In: Yost WA, Popper AN, Fay R, York N (eds) Auditory perception of sound sources. Springer, New York
- CARLYON RP, CUSACK R, FOXTON JM, ROBERTSON IH (2001) Effects of attention and unilateral neglect on auditory stream segregation. *J Exp Psychol: Hum Percept Perform* 27:115–127
- CHISTOVITCH LA (1959) Perception of time interval between two short sound pulses. *Soviet Physics: Acoustics* 5:480–484
- CHURCH RM, GETTY DJ, LERNER ND (1976) Duration discrimination by rats. *J Exp Psychol: Anim Behav Process* 2:303–312
- CREELMAN CD (1964) Human discrimination of auditory duration. In: Swets JA (ed) Signal detection and recognition by human observers. Wiley, New York, pp 265–290
- CUSACK R, ROBERTS B (1999) Effects of similarity in bandwidth on the auditory sequential streaming of two-tone complexes. *Perception* 28:1281–1289
- CUSACK R, DEEKS J, AIKMAN G, CARLYON RP (2004) Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *J Exp Psychol: Hum Percept Perform* 30:643–656
- DARWIN CJ, CARLYON RP (1995a) Auditory organization and the formation of perceptual streams. In: Moore BCJ (ed) Handbook of perception and cognition, volume 6, Hearing. Academic, San Diego
- DARWIN CJ, CARLYON RP (1995b) Auditory grouping. In: Moore BCJ (ed) Hearing. Academic, London, pp 387–424
- DIVENYI PL, DANNER WF (1977) Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra. *Percept Psychophys* 21:125–142
- DOWLING WJ (1973) The perception of interleaved melodies. *Cognitive Psychol* 5:322–337
- DRAKE C, BOTTE MC (1993) Tempo sensitivity in auditory sequences: evidence for a multiple-look model. *Percept Psychophys* 54:277–286
- DURLACH NI, MASON CR, SHINN-CUNNINGHAM BG, ARBOGAST TL, COLBURN HS, KIDD CJ (2003) Informational masking: counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *J Acoust Soc Am* 114:368–379
- ELHILALI M, MA L, MICHEYL C, OXENHAM AJ, SHAMMA SA (2009) Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron* 61:317–329
- FAY RR (1998) Auditory stream segregation in goldfish (*Carassius auratus*). *Hear Res* 120:69–76
- FAY RR (2000) Spectral contrasts underlying auditory stream segregation in goldfish (*Carassius auratus*). *J Assoc Res Otolaryngol* 1:120–128
- FAY RR (2008) Sound source perception and stream segregation in nonhuman vertebrate animals. In: Yost WA, Popper AN, Fay RR (eds) Auditory perception of sound sources. Springer, New York, pp 307–323
- FIELLER EC (1957) Tests for rank correlation coefficients: I. *Biometrika* 44:470–481
- FISHMAN YI, RESER DH, AREZZO JC, STEINSCHNEIDER M (2001) Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hear Res* 151:167–187
- FISHMAN YI, AREZZO JC, STEINSCHNEIDER M (2004) Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. *J Acoust Soc Am* 116:1656–1670
- FITZGIBBONS PJ, GORDON-SALANT S (1994) Age effects on measures of auditory duration discrimination. *J Speech Hear Res* 37:662–670
- FITZGIBBONS PJ, GORDON-SALANT S (1995) Age effects on duration discrimination with simple and complex stimuli. *J Acoust Soc Am* 98:3140–3145
- FITZGIBBONS PJ, GORDON-SALANT S, BARRETT J (2007) Age-related differences in discrimination of an interval separating onsets of successive tone bursts as a function of interval duration. *J Acoust Soc Am* 122:458–466
- FRENCH-ST GEORGE M, BREGMAN AS (1989) Role of predictability of sequence in auditory stream segregation. *Percept Psychophys* 46:384–386
- GAUDRAIN E, GRIMAULT N, HEALY EW, BERA JC (2007) Effect of spectral smearing on the perceptual segregation of vowel sequences. *Hear Res* 231:32–41
- GAUDRAIN E, GRIMAULT N, HEALY EW, BERA JC (2008) Streaming of vowel sequences based on fundamental frequency in a cochlear-implant simulation. *J Acoust Soc Am* 124:3076–3087
- GOCKEL H, CARLYON RP, MICHEYL C (1999) Context dependence of fundamental-frequency discrimination: lateralized temporal fringes. *J Acoust Soc Am* 106:3553–3563
- GREEN DM (1995) Maximum-likelihood procedures and the inattentive observer. *J Acoust Soc Am* 97:3749–3760
- GREEN DM, SWETS JA (1966) Signal detection theory and psychophysics. Krieger, New York
- GRIMAULT N, MICHEYL C, CARLYON RP, ARTHAUD P, COLLET L (2000) Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency. *J Acoust Soc Am* 108:263–271
- GRIMAULT N, MICHEYL C, CARLYON RP, ARTHAUD P, COLLET L (2001) Perceptual auditory stream segregation of sequences of complex

- sounds in subjects with normal and impaired hearing. *Br J Audiol* 35:173–182
- GRIMAULT N, BACON SP, MICHEYL C (2002) Auditory stream segregation on the basis of amplitude-modulation rate. *J Acoust Soc Am* 111:1340–1348
- HARTMANN WM, JOHNSON D (1991) Stream segregation and peripheral channeling. *Music Percept* 9:155–184
- HIENZ RD, SINNOTT JM, SACHS MB (1980) Auditory intensity discrimination in blackbirds and pigeons. *J Comp Physiol Psychol* 94:993–1002
- HIRSH IJ, MONAHAN CB, GRANT KW, SINGH PG (1990) Studies in auditory timing: 1. Simple patterns. *Percept Psychophys* 47:215–226
- HUANG R, RICHARDS VM (2006) Coherence detection: effects of frequency, frequency uncertainty, and onset/offset delays. *J Acoust Soc Am* 119:2298–2304
- HULSE SH, MACDOUGALL-SHACKLETON SA, WISNIEWSKI AB (1997) Auditory scene analysis by songbirds: stream segregation of birdsong by European starlings (*Sturnus vulgaris*). *J Comp Psychol* 111:3–13
- ITATANI N, KLUMP GM (2009) Auditory streaming of amplitude-modulated sounds in the songbird forebrain. *J Neurophysiol* 101:3212–3225
- IZUMI A (2002) Auditory stream segregation in Japanese monkeys. *Cognition* 82:B113–B122
- KANWAL JS, MEDVEDEV AV, MICHEYL C (2003) Neurodynamics for auditory stream segregation: tracking sounds in the mustached bat's natural environment. *Network* 14:413–435
- KIDD GJ, MASON CR, DELIWALA PS, WOODS WS, COLBURN HS (1994) Reducing informational masking by sound segregation. *J Acoust Soc Am* 95:3475–3480
- KIDD GJ, MASON CR, ARBOGAST TL (2002) Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns. *J Acoust Soc Am* 111:1367–1376
- KIDD GJ, MASON CR, RICHARDS VM (2003) Multiple bursts, multiple looks, and stream coherence in the release from informational masking. *J Acoust Soc Am* 114:2835–2845
- LEVITT H (1971) Transformed up-down methods in psychoacoustics. *J Acoust Soc Am* 49:467–477
- LOGOTHETIS NK, SCHALL JD (1989) Neuronal correlates of subjective visual perception. *Science* 245:761–763
- MA L, MICHEYL C, YIN P, OXENHAM AJ, SHAMMA S (2010) Behavioral measures of auditory streaming in ferrets. *J Comp Psychol* (in press)
- MACDOUGALL-SHACKLETON SA, HULSE SH, GENTNER TQ, WHITE W (1998) Auditory scene analysis by European starlings (*Sturnus vulgaris*): perceptual segregation of tone sequences. *J Acoust Soc Am* 103:3581–3587
- MACMILLAN NA, CREELMAN CD (2005) *Detection theory: a user's guide*. Erlbaum, Mahwah
- MAIER EH, KLUMP GM (1990) Auditory duration discrimination in the European starling (*Sturnus vulgaris*). *J Acoust Soc Am* 88:616–621
- MCCABE S, DENHAM MJ (1997) A model of auditory streaming. *J Acoust Soc Am* 101:1611–1621
- MICHEYL C, CARIYON RP (1998) Effects of temporal fringes on fundamental-frequency discrimination. *J Acoust Soc Am* 104:3006–3018
- MICHEYL C, TIAN B, CARIYON RP, RAUSCHECKER JP (2005A) Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron* 48:139–148
- MICHEYL C, CARIYON RP, CUSACK R, MOORE BCJ (2005B) Performance measures of auditory organization. In: Pressnitzer D, de Cheveigné A, McAdams S, Collet L (eds) *Auditory signal processing: physiology, psychoacoustics, and models*. Springer, New York, pp 203–211
- MICHEYL C, SHAMMA S, OXENHAM AJ (2007A) Hearing out repeating elements in randomly varying multitone sequences: a case of streaming? In: Kollmeier B, Klump G, Hohmann V, Lange-mann U, Mauermann M, Uppenkamp S, Verhey J (eds) *Hearing—from basic research to applications*. Springer, Berlin, pp 267–274
- MICHEYL C, CARIYON RP, GUTSCHALK A, MELCHER JR, OXENHAM AJ, RAUSCHECKER JP, TIAN B, COURTENAY WILSON E (2007B) The role of auditory cortex in the formation of auditory streams. *Hear Res* 229:116–131
- MICHEYL C, OXENHAM AJ (2010) Pitch, harmonicity, and concurrent sound segregation: psychoacoustical and neurophysiological findings. *Hearing Res* 266(1–2):36–51
- MILLER GA, HEISE GA (1950) The trill threshold. *J Acoust Soc Am* 22:637–638
- MOORE BCJ, GOCKEL H (2002) Factors influencing sequential stream segregation. *Acta Acustica United with Acustica* 88:320–333
- OKADA M, KASHINO M (2008) Temporal jitter affects perceptual transitions in auditory streaming. *Proceedings of the 31st MidWinter Research Meeting of the Association for Research in Otolaryngology*
- OXENHAM AJ, DAU T (2001) Modulation detection interference: effects of concurrent and sequential streaming. *J Acoust Soc Am* 110:402–408
- OXENHAM AJ, FLIGOR BJ, MASON CR, KIDD G Jr (2003) Informational masking and musical training. *J Acoust Soc Am* 114:1543–1549
- PARKER AJ, NEWSOME WT (1998) Sense and the single neuron: probing the physiology of perception. *Annu Rev Neurosci* 21:227–277
- PHILLIPS SL, GORDON-SALANT S, FITZGIBBONS PJ, YENI-KOMSHIAN GH (1994) Auditory duration discrimination in young and elderly listeners with normal hearing. *J Am Acad Audiol* 5:210–215
- PRESSNITZER D (2008) The build-up of streaming adapts to sequence duration. *Abstracts of the 31st Annual Midwinter Research Meeting of the Association for Research in Otolaryngology* 141–142
- PRESSNITZER D, SAYLES M, MICHEYL C, WINTER IM (2008) Perceptual organization of sound begins in the auditory periphery. *Curr Biol* 18:1124–1128
- RAMMSAYER TH (1994) Effects of practice and signal energy on duration discrimination of brief auditory intervals. *Percept Psychophys* 55:454–464
- RAMMSAYER TH, LIMA SD (1991) Duration discrimination of filled and empty auditory intervals: cognitive and perceptual factors. *Percept Psychophys* 50:565–574
- ROBERTS B, GLASBERG BR, MOORE BC (2002) Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. *J Acoust Soc Am* 112:2074–2085
- ROBERTS B, GLASBERG BR, MOORE BC (2008) Effects of the build-up and resetting of auditory stream segregation on temporal discrimination. *J Exp Psychol: Hum Percept Perform* 34:992–1006
- ROGERS WL, BREGMAN AS (1993) An experimental evaluation of three theories of auditory stream segregation. *Percept Psychophys* 53:179–189
- SCHUL J, SHERIDAN RA (2006) Auditory stream segregation in an insect. *Neuroscience* 138:1–4
- SHAMMA SA, MICHEYL C (2010) Behind the scenes of auditory perception. *Curr Opin Neurobiol* 20:361–366
- SINNOTT JM, OWREN MJ, PETERSEN MR (1987) Auditory duration discrimination in Old World monkeys (*Macaca, Cercopithecus*) and humans. *J Acoust Soc Am* 82:465–470
- SNYDER JS, ALAIN C (2007) Toward a neurophysiological theory of auditory stream segregation. *Psychol Bull* 133:780–799

- SUSSMAN E, RITTER W, VAUGHAN HGJ (1999) An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiol* 36:22–34
- SWETS JA, SHIPLEY EF, McKEY MJ, GREEN DM (1959) Multiple observations of signals in noise. *J Acoust Soc Am* 31:514–521
- VAN NOORDEN LPAS (1975) Temporal coherence in the perception of tone sequences. Unpublished doctoral dissertation. University of Technology, Eindhoven
- VIEMEISTER NF, SCHLAUCH RS (1994) Issues in infant psychoacoustics. In: Wernerand LA, Rubel EW (eds) *Developmental psychoacoustics*. American Psychological Association, Washington, pp 191–209
- VLIEGEN J, OXENHAM AJ (1999) Sequential stream segregation in the absence of spectral cues. *J Acoust Soc Am* 105:339–346
- VLIEGEN J, MOORE BC, OXENHAM AJ (1999) The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. *J Acoust Soc Am* 106:938–945
- WIGHTMAN E, ALLEN P (1994) Individual differences in auditory capacity among preschool children. In: Wernerand LA, Rubel EW (eds) *Developmental psychoacoustics*. American Psychological Association, Washington, pp 113–133
- WIGHTMAN E, ALLEN P, DOLAN T, KISTLER D, JAMIESON D (1989) Temporal resolution in children. *Child Dev* 60:611–624
- WRIGHT BA, DAI H (1994) Detection of unexpected tones with short and long durations. *J Acoust Soc Am* 95:931–938
- WRIGHT BA, BUONOMANO DV, MAHNCKE HW, MERZENICH MM (1997) Learning and generalization of auditory temporal-interval discrimination in humans. *J Neurosci* 17:3956–3963