Journal of Experimental Psychology:
Human Perception and Performance
1999, Vol. 25, No. 6, 1742–1759

Perceptual Organization of Complex Auditory Sequences: Effect of Number of Simultaneous Subsequences and Frequency Separation

Renaud Brochard, Carolyn Drake, and
Marie-Claire Botte
Université René Descartes

Stephen McAdams
Université René Descartes and Institut de Recherche et de
Coordination Acoustique/Musique

Previous findings on streaming are generalized to sequences composed of more than 2 subsequences. A new paradigm identified whether listeners perceive complex sequences as a single unit (integrative listening) or segregate them into 2 (or more) perceptual units (stream segregation). Listeners heard 2 complex sequences, each composed of 1, 2, 3, or 4 subsequences. Their task was to detect a temporal irregularity within 1 subsequence. In Experiment 1, the smallest frequency separation under which listeners were able to focus on 1 subsequence was unaffected by the number of co-occurring subsequences; nonfocused sounds were not perceptually organized into streams. In Experiment 2, detection improved progressively, not abruptly, as the frequency separation between subsequences increased from 0.25 to 6 auditory filters. The authors propose a model of perceptual organization of complex auditory sequences.

When listening to a complex sound environment composed of multiple sound sources, listeners must organize incoming information in such a way as to separate the sound mixture into distinct sources and perceptually link, over time, events that belong to the same source (stream segregation and sequential grouping, respectively). These processes are essential if listeners are to be able to concentrate on one musical voice within a complex symphony or on a particular conversation in a crowded room (the well-known "cocktail party problem" described by Cherry in 1953). In this article, we examine how listeners perceptually organize complex sequences of two or more concurrent subsequences. Manipulation of structural characteristics of complex sequences (such as the frequency and temporal separation between adjacent subsequences) allowed us to study how these features influence perceptual organization and, thus, to deduce the underlying processes.

The intriguing characteristic of the perceptual organization of complex auditory sequences is that the perceived structure varies considerably depending on certain structural features of the sequences. Take, as an example, the stimuli traditionally used in this type of experiment. Figure 1A shows a complex auditory sequence composed of two isochronous subsequences of pure tones. Each subsequence is defined by a specific frequency–tempo (interonset interval [IOI]) combination. An impressive body of research (Bregman & Campbell, 1971; Jones, 1976; van Noorden, 1975; see Bregman, 1990, for a review) has established that the perceived structure of these sequences depends to a large extent on the relationship between the frequency and tempo of the two subsequences. One major determining characteristic concerns the difference in frequency between the two subsequences (see Figure 1B): If the frequency difference is small, listeners perceive a single sequence composed of the alternation of the two subsequences (by a process of serial integration). However, if the frequency difference exceeds a certain limit (van Noorden's fission boundary), the percept changes considerably: Listeners sometimes report hearing one of the two subsequences as a figure, with the other subsequence making up the background (Bregman, 1990). This organizational process is known as *stream segregation* (Bregman & Campbell, 1971). Usually, listeners can choose to focus attention on one of the subsequences or the other, creating focused and nonfocused streams. In addition to the influence of frequency separation, the perceptual organization of the complex sequence also depends on the *temporal structure* of the subsequences (Bregman, 1990; Jones, 1976; Jones, Maser, & Kidd, 1978; van Noorden, 1975). Figure 1C shows the *interaction* between frequency and temporal relations: At faster rates, streaming becomes more likely, even with relatively small frequency separations (Jones & McCallum, 1987; van Noorden, 1975).

Figure 1C presents the three perceptual zones described by van Noorden (1975) as a function of the frequency and temporal separation between the two subsequences. Below the *fission boundary,* listeners are unable to follow just one

Renaud Brochard, Carolyn Drake, and Marie-Claire Botte, Laboratoire de Psychologie Expérimentale, Centre National de la Recherche Scientifique, Université René Descartes, Paris, France; Stephen McAdams, Laboratoire de Psychologie Expérimentale, Centre National de la Recherche Scientifique, Université René Descartes, and Institut de Recherche et de Coordination Acoustique/Musique, Paris, France.

This work was completed in partial requirement for Renaud Brochard's doctoral dissertation, begun under the direction of Marie-Claire Botte.

Correspondence concerning this article should be addressed to Renaud Brochard, who is now at the Psychology Department, University of Keele, Newcastle-Under-Lyme, Staffordshire, ST5 5BG England, or to Carolyn Drake, Laboratoire de Psychologie Expérimentale, CNRS UMR 8581, Université René Descartes, 28 rue Serpente, 75006 Paris, France. Electronic mail may be sent to Renaud Brochard at r.brochard@keele.ac.uk or Carolyn Drake at drake@idf.ext.jussieu.fr.

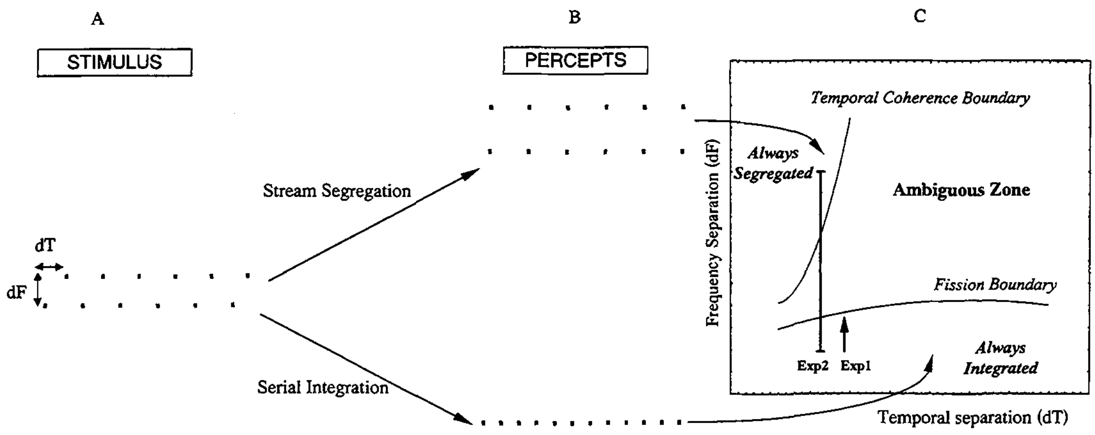

Figure 1. A: Sequence composed of tones of two frequencies typically used in streaming experiments. B: Two possible percepts into one or two streams. C: Schematic presentation of the perceptual results of the interaction between frequency and temporal separations (after van Noorden, 1975). Exp = experiment.

subsequence; rather, they integrate all of the sounds into a single stream. This perceptual limit depends mainly on the frequency separation between the alternating tones (less than 3–4 semitones) which, according to Bregman (1990), reflects peripheral processing constraints because this frequency separation corresponds approximately to the width of one auditory filter (roughly speaking, the zone corresponding to the frequency selectivity of sensory cells in the cochlea; see Glasberg & Moore, 1990, for more precise psychoacoustic definitions in terms of masking). Above the *temporal coherence boundary,* listeners always hear two streams. This perceptual limit was first described by Miller and Heise (1950) as the point at which sequential integration of all of the tones into one percept is no longer possible (trill threshold). This boundary limiting the integration of the entire sequence depends on both the frequency separation of the alternating tones and the rate at which the events occur: Larger frequency separations require a slower presentation rate to maintain a serially integrated perception of the sequence (i.e., one temporally coherent stream). Jones (1976) and van Noorden (1975), as well as Anstis and Saida (1985), have proposed that this phenomenon can be explained by "pitch motion" processes. These processes appear to function efficiently under a limit of about 0.15 semitones per millisecond. For Jones (1976), the *selective attention boundary* (van Noorden's fission boundary) also reflects a pitch motion limit, but at a much lower rate of about 0.006 semitones per millisecond. Sequences that fall between these two boundaries (in the *ambiguous zone*) can be perceived in either fashion (one or two streams) depending on the attentional requirements of the experimental task. Attention can be manipulated experimentally by inducing listeners to focus (or not) on particular physical characteristics within the sequence by providing a cue (Botte, Drake, Brochard, & McAdams, 1997; Davidson, Power, & Michie, 1987) or by creating expectations (Dowling, 1973; Dowling, Lung, & Herrbold, 1987; Jones & Boltz, 1989), leading to the sequence being perceptually organized in a particular fashion.

The main aim of this article is to generalize these classic results that have been obtained with two subsequences to more complex sequences composed of two, three, or four subsequences. We address three questions. The first concerns the size of frequency separation necessary for stream segregation when there are three (or more) subsequences and, thus, two (or more) frequency separations involved. Is the same size of frequency separation required between each of the pairs of subsequences as with simpler sequences, or is a larger frequency separation necessary because it must be separated from more potential streams? The second question concerns the perceptual organization of nonfocused subsequences. In the traditional case of two subsequences, there are just the focused and nonfocused streams, but in the more complex case of three subsequences, there are the focused stream and the others: Are the others segregated into separate streams (thus creating three perceptual units or streams), or are the events in the other subsequences mixed into a single perceptual unit? The third question concerns the interrelationship among the three perceptual zones described above (below the fission boundary, in the ambiguous zone, and above the temporal coherence boundary): Does perception change abruptly and qualitatively among the three zones, reflecting distinct sets of processes, or is there a gradual passage from one to the other reflecting a continuation in underlying processes?

A second aim of these studies was to investigate the role of skill in this perceptual organization. Musicians have had considerable experience in listening analytically to complex sequences. Do they hear sequences in a different way from novice listeners (nonmusicians)? Experimental evidence suggests that musicians have heightened sensitivity to temporal aspects of a sequence relative to nonmusicians; musicians are able to detect smaller tempo changes (Drake & Botte, 1993) and temporal irregularities (Jones, Jagacinski, Yee, Floyd, & Klapp, 1995) than nonmusicians. Evidence in the literature (e.g., Davidson et al., 1987; Smith, Hausfeld, Power, & Gorta, 1982) suggests that musicians perceptually organize complex sequences in a different way from nonmusicians. However, Jones et al. (1995) found no evidence of increased attentional flexibility in musicians as compared with nonmusicians; whereas musicians were more able to detect a temporal jitter, they were not more likely to change from selective to integrative attending in accordance with the optimum strategy. Here we tested the further possibility that improvements with skill may involve shifts of the fission boundary; stream segregation may occur with smaller frequency separations for expert listeners.

A new paradigm has been developed to address these issues. The aim was to create an objective measure of stream segregation, in contrast to previous tasks that have tended to be rather subjective, with participants required to recognize a particular rhythm (e.g., van Noorden, 1975), choose between possible organizations (e.g., Davidson et al., 1987), or estimate degree of segregation on a numerical scale (Rogers & Bregman, 1993). Our experimental rationale involved asking listeners to detect a small temporal irregularity hidden within a complex sequence so that the task could be accomplished only if listeners perceptually organized the sequence into streams.

An example of one trial is presented in Figure 2. Participants heard two complex sequences, each composed of two to four subsequences. They were asked to detect a temporal irregularity embedded in one subsequence (the target subsequence) of one complex sequence. Their task was to indicate whether the temporal irregularity was in the first or second complex sequence. Attention was directed toward the target subsequence by preceding the complex sequences with a simple cue sequence of the same frequency–tempo combination as the target subsequence. The principle behind these experiments revolved around the fact that the ease with which listeners can detect a temporal irregularity depends on the tempo of the sequence: Temporal sensitivity is best at intermediate tempi (200-ms to 800-ms IOI) and decreases considerably at faster and slower rates (Drake & Botte, 1993; Friberg & Sundberg, 1995; Hibi, 1983). Thus, if the complex sequence is perceptually segregated into streams, the temporal irregularity has to be detected in relation to the tempo of the focused target subsequence. The size of the temporal irregularity was fixed in such a way as to be well above threshold and thus easily detectable in these conditions. However, if the sequence is not perceptually segregated into streams, the reference tempo is a combination of the tempo of all of the subsequences (i.e., an irregular, faster tempo).

The complex temporal framework of the entire sequence is shown at the bottom of Figure 2. The tempi of the subsequences (IOIs of 700, 500, 400, and 300 ms) were in inharmonic ratios (e.g., simple multiples of two or three) to prevent listeners from extracting a hierarchical rhythmic structure (e.g., binary and ternary rhythms). The size of temporal irregularity used here was below the threshold associated with this framework and thus could not be detected. This paradigm is similar to the one used by Jones et al. (1995), although they used a fixed-size temporal irregularity and manipulated the attentional set by changing the instructions.

Our experimental rationale is valid only if we can demonstrate two facts. First, the probability of detecting the

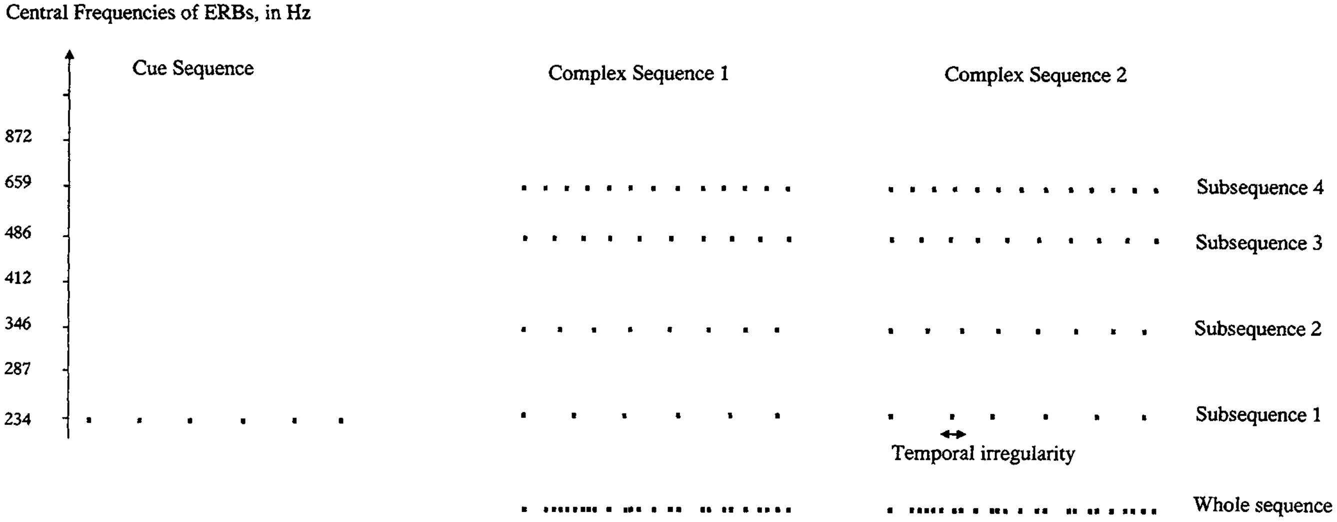

Figure 2. Experiment 1: Each trial consisted of a single cue sequence followed by two complex sequences composed of 1, 2, 3, or 4 subsequences. In this example, the 4 subsequences are presented (Subsequence 1, 700 ms; Subsequence 2, 500 ms; Subsequence 3, 400 ms; and Subsequence 4, 300 ms), and Subsequence 1 is cued. The arrow shows an exaggerated example of a temporal irregularity in the cued subsequence. Higher frequency subsequences were progressively added and cued during the experiment. See text for details of the frequency separation scale. ERBs = equivalent rectangular bandwidths.

temporal irregularity must be similar in all four subsequences. Control Experiment 1 verified that this was the case by measuring temporal irregularity detection thresholds in simple sequences (one subsequence) varying in tempo from 300–700 ms IOIs. Second, it must be demonstrated that the temporal irregularity cannot be detected if listeners do not perceptually organize the sequence into streams, using a global listening strategy instead. Control Experiment 2 verified this point by measuring percentage correct detection of a temporal irregularity under conditions in which stream segregation should be impossible to show that the temporal irregularity is not detected in relation to the sequence as a whole.

We present two main experiments using this rationale in different experimental paradigms to provide converging evidence for the processes involved in the perceptual organization of complex sequences. In Experiment 1, we first established the minimum frequency separation (MFS) between two adjacent subsequences under which each listener was able to focus attention on each subsequence and thus detect the temporal irregularity in the focused subsequence. This was done by gradually increasing the frequency separation by steps of one auditory filter bandwidth (expressed in equivalent rectangular bandwidths [ERBs], a commonly used mathematical method for characterizing the width of auditory filters; Glasberg & Moore, 1990) until the listener was able to focus efficiently enough on a particular subsequence to allow estimation of a threshold by an adaptive two-alternative forced-choice method (no convergence is obtained if responses are random). The MFS was established in this way for each subsequence in relation to all other adjacent subsequences for complex sequences composed of two, three, or four subsequences. The question of the perceptual effort required when one is processing concurrent subsequences was addressed by comparing the values of the thresholds obtained in these borderline conditions. If all of the unattended subsequences are processed within the same perceptual unit, an increasing number of distracting subsequences should not alter performance, so we expected constant thresholds. Alternatively, if each unattended subsequence is processed separately, temporal detection performance should decrease as a result of increased processing load.

Experiment 2 investigated the improvement of selective attending as the frequency separation between adjacent subsequences increased above the extreme focusing conditions of Experiment 1. Most theories of stream segregation (Bregman, 1990; Jones, 1976; van Noorden, 1975) predict that focusing on one stream should become easier as stimuli are situated at different points along a continuum: As the frequency separation increases, the stimuli are situated initially below the fission boundary, then within the intermediate zone, and finally above the temporal integration boundary (see Figure 1C). The possible percepts should thus change from a situation in which it is impossible to focus on one subsequence (integrated listening), through a state in which focusing either on individual subsequences or on their integration into a single stream is possible, to a state of obligatory segregation and focusing on individual streams. The idea of zones presented by previous authors suggests three distinct perceptual states. This may be a useful descriptive tool, but we suggest, rather, that there is a gradual passage from one state to another, with the probability of successfully focusing on one stream gradually increasing with an increased frequency separation. We tested this hypothesis by measuring the ability of listeners to detect a fixed-size temporal irregularity in each of the three subsequences of a complex sequence as the frequency separation passed from one extreme to the other. We also predicted that the probability of successfully focusing on one particular stream would depend on certain physical characteristics of the sequence (position of the subsequence within the complex sequence) and on the focusing abilities of the listener. We therefore measured the probability of successfully focusing on one stream for (a) subsequences with the highest, lowest, and intermediate frequencies; (b) subsequences surrounded by subsequences of different tempo–frequency combinations; and (c) musician and nonmusician listeners.

Control Experiment 1: Detection Unaffected by Subsequence Tempo

Experiments 1 and 2 involved the use of complex sequences composed of at least two subsequences, each defined by a specific tempo and frequency. The experimental rationale was based on the fact that detection of a temporal irregularity is equivalent in all subsequences. As a result of the long experimental procedures, not all tempo–frequency combinations could be tested in the main experiments. Previous studies (see Botte et al., 1997) have indicated no differences in detection for the frequency range used here. However, it remains to be established that a small temporal irregularity (in sequences presented on their own) is equally detectable in each of the four tempi used (IOIs of 700, 500, 400, and 300 ms). These tempi were chosen because they fall into the range of optimal temporal sensitivity observed previously (Drake & Botte, 1993; Friberg & Sundberg, 1995; Hibi, 1983; Hirsh, Monahan, Grant, & Singh, 1990). Weber's law should apply within this range, and so we predicted similar thresholds for the four tempi.

Method

Stimuli. Each trial consisted of the successive presentation of two isochronous sequences composed of 234 Hz tones with IOIs of 300, 400, 500, or 700 ms. The total duration of each sequence was fixed at 3,650 ms, and so the number of tones in the sequence varied with the tempo (6, 8, 10, and 13 tones, respectively). A temporal irregularity was created by advancing or delaying, in relation to regularity, the onset of one tone in one of the sequences, near the beginning, middle, or end of the sequence. A silence of 1,700 ms separated the two sequences. Tones had a duration of 50 ms (including 10-ms onset and offset ramps) and were presented to both ears at 70 dB SPL. All of the subsequences began at the same time, to avoid listeners' attention being drawn toward the frequency region of the tone heard first.

Procedure. Participants heard two sequences. Their task was to indicate whether the temporal irregularity was in the first or second sequence. Thresholds were calculated with an adaptive method (1 up–4 down) leading to a theoretical 84.1% detection level (Levitt, 1971). Four successive correct answers resulted in a decrease of 1%; one incorrect answer led to an increase of 1%. The threshold measurement stopped after six reversals (changes in direction), and detection thresholds were calculated by averaging over the last four reversals. The initial value of the irregularity was 5% of the reference tempo. This value was adopted because pilot studies indicated that thresholds were close to this value, thus increasing accuracy and reducing the length of experimental sessions. The first five responses were excluded from the calculation. On average, 35 trials were required to reach threshold criteria. Each measurement was repeated three times.

Apparatus. Sequences were generated by a digital signal processing card (OROS) and controlled by a personal computer. Participants sat in a soundproof room and listened to sequences through headphones (TDH 49). A programmable attenuator (Charybdis D) controlled sound levels. Participants gave their response by pressing one of two buttons, and they received visual feedback indicating the correctness of their response.

Participants. Ten listeners took part in this experiment. They all had normal hearing and were paid for their participation. All were undergraduate psychology students at René Descartes University who had not previously participated in psychoacoustic experiments.

Results and Discussion

A repeated measures analysis of variance (ANOVA) on detection thresholds with tempo (four levels) and repetitions (three levels) as independent variables yielded no significant main effects or their interaction. Detection rates, averaged over repetitions, were almost the same for all of the tempi (5.8%, 5.9%, 4.3%, and 5.3% for IOIs of 300, 400, 500, and 700 ms, respectively). These values are in good agreement with those generally found for irregularity detection in isochronous sequences (Friberg & Sundberg, 1995; Hibi, 1983; Hirsh et al., 1990). Thresholds are relatively steady and range from 4% to 7%, depending on the method used. In the tempo region studied here (300-ms to 700-ms IOIs), Weber's law seems to describe correctly temporal detection sensitivity. It corresponds to the classical region (200–800 ms) of optimum temporal sensitivity (Drake & Botte, 1993; Fraisse, 1967; Jones & Boltz, 1989). These four tempi were therefore used in the main experiments.

Control Experiment 2: Temporal Irregularity Undetectable Under Global Listening

The experimental rationale of the two main experiments was based on the assumption that listeners are unable to detect temporal irregularity if they have not perceptually organized the sequence into streams; they use a global listening strategy (by a process of serial integration). Temporal irregularity should be undetectable in such conditions as a result of the much larger Weber fractions required because the temporal irregularity must be detected in relation to the rapid, irregular rhythm presented at the bottom of Figure 2. To test this assumption, we measured listeners' detection of a temporal irregularity under conditions in which stream segregation should be impossible (small frequency separation and without a cue) to verify that the temporal irregularity was not detected in relation to the sequence as a whole. Thus, in this control experiment, we compared the detection performance for the irregularity in the presence or absence of a cue sequence, in conditions in which stream segregation was easy (large frequency separation) or hard (small frequency separation). We predicted that performance would be near chance level when streaming was impossible, that is, when the frequency separation was small or when the target subsequence was not cued.

Method

Stimuli. We presented complex sequences composed of three subsequences. The three IOIs of the subsequences were 700 ms (Subsequence 1 [S1]), 500 ms (Subsequence 2 [S2]), and 300 ms (Subsequence 3 [S3]). The number of tones in each subsequence varied so as to provide a fixed total duration of 3,650 ms (6, 8, and 13 tones, respectively). The frequency separation between these subsequences was varied on a psychoacoustically derived scale thought to correspond loosely to the position of maximal stimulation of a pure tone along the basilar membrane. This scale is expressed in units related to the estimated bandwidth of auditory filters. The ERB (in Hz) is calculated according to the following formula: $24.7 \times (4.37F + 1)$, where F is the central frequency expressed in kHz (Glasberg & Moore, 1990). In our stimuli, frequencies were separated by 1 or 6 ERBs. A local temporal irregularity was always placed in Subsequence 2 (target subsequence), but the frequency varied from trial to trial, so the target could be the highest, middle, or lowest subsequence.

Procedure. In a two-interval forced-choice paradigm, we presented two complex sequences each composed of three subsequences. Listeners were required to indicate whether the irregularity was in the first or second complex sequence. The size of the irregularity was well above detection threshold (75 ms: 15% of the 500-ms target tempo). The frequency position (high, middle, or low) of the target stream and the amount of frequency separation between the subsequences were randomized over trials. For half of the trials, the two complex sequences were preceded by a cue sequence corresponding in frequency and tempo to the target stream (cued condition). For the other half, this cue sequence was replaced by a silence of the same duration (noncued condition). In all, listeners completed 120 trials (2 cues × 3 positions × 2 frequency separations × 10 repetitions).

Apparatus. The apparatus was the same as in Control Experiment 1.

Participants. Thirty-one listeners took part in this experiment. They all had normal hearing and were paid for their participation. All were undergraduate psychology students at René Descartes University who had not previously participated in psychoacoustic experiments.

Results and Discussion

An ANOVA on condition (cued vs. noncued), frequency separation (1 vs. 6 ERBs), and frequency position (high vs. middle vs. low) showed an effect of frequency separation, $F(1, 30) = 11.66, p < .002$, and cuing condition, $F(1, 30) = 8.96, p < .005$, but not an effect of frequency position. There was a significant interaction between condition and position, $F(1, 30) = 7.42, p < .01$. Detection performance, averaged over frequency position, is summarized in Table 1. Listeners

Table 1
Control Experiment 2: Correct Detection Percentages for the Easy and Hard Focusing Conditions With and Without a Cue Sequence

Condition	6 ERBs (easy streaming)	1 ERB (hard streaming)
With cue	90.4	60.2
Without cue	62.4	59.1

Note. ERB = equivalent rectangular bandwidth.

were unable to detect the temporal irregularity with either a small frequency separation or without the presence of a cue sequence. Detection was possible only with the wide frequency separation in the presence of the cue. As predicted, listeners were unable to detect the temporal irregularity under conditions that made stream segregation very difficult, thus implying a global listening strategy. The second experimental assumption was therefore confirmed. The results of these two control experiments thus provide evidence in support of the experimental rationale used in the two main experiments.

Experiment 1: Effect of Number of Co-Occurring Subsequences

Experiment 1 investigated how the process of stream segregation is influenced by the surrounding context. The large amount of research dealing with stream segregation has concentrated primarily on sequences containing only two subsequences and does not predict the effect of a larger number of distracting subsequences. Here we compared listeners' ability to detect temporal irregularities in conditions of increasing mixture complexity. Complexity was manipulated by varying the number of potential streams (from one to four), which we call subsequences. The term *streams* is reserved for the perceived grouping of similar tones into a perceptual unit. Does it become more difficult to focus on one subsequence when it is embedded in an increasingly complex mixture? Are nonfocused subsequences processed separately with as many perceptual units as there are potential streams, or are they processed as a single perceptual unit?

To address these issues, we needed an indication of the ease of stream segregation. This was done by obtaining a measure of the MFS between adjacent subsequences necessary for listeners to focus efficiently on a particular subsequence. This is equivalent to measuring the position of van Noorden's fission boundary. We investigated how this MFS varied under different experimental conditions: We expected it to be larger for harder processing conditions and, thus, larger for mixtures containing more subsequences. The MFS was established by gradually increasing the frequency separation between adjacent subsequences by steps of one auditory filter bandwidth. Listeners were considered to have perceptually organized the sequence into streams once they were able to focus on the cued subsequence and to detect the temporal irregularity embedded within it. Irregularity detection thresholds were measured under these borderline conditions.

Method

Stimuli. Sequences were composed of one, two, three, or four simultaneous subsequences of pure tones that were defined by tempo and frequency (see Figure 2). S1 had the lowest frequency (234 Hz) and the slowest tempo (700-ms IOI) and remained constant throughout the experiment. Other higher and faster subsequences were added at frequencies above S1.

The tempi of S2–S4 were fixed (S2, 500-ms IOI; S3, 400-ms IOI; and S4, 300-ms IOI). Each sequence lasted 3,650 ms, so the number of tones varied for each subsequence (S1 = 6 tones, S2 = 8 tones, S3 = 10 tones, and S4 = 13 tones). All of the subsequences began at the same time to avoid listeners' attention being drawn toward the frequency region of the tone heard first.

The frequencies of S2 to S4 were varied; they corresponded to multiples of a 1-ERB frequency separation above 234 Hz. A local temporal irregularity was created by advancing or delaying the onset of one of the tones in one of the subsequences in relation to regularity (see Figure 2). As a means of increasing stimulus uncertainty and thus enhancing the need to maintain focus throughout the sequence, the irregularity could occur in one of three positions, near the beginning, middle, or end of the sequence (Tones 3, 4, and 5 for S1; tones 3, 5, and 6 for S2; Tones 4, 5, and 9 for S3; and Tones 5, 7, and 10 for S4). Tone durations were 50 ms (including 10-ms onset and offset ramps) and were presented to both ears at the same level (70 phons; see Scharf & Houstma, 1986). The two complex sequences were preceded by a cue sequence that had the same tempo and frequency as the subsequence containing the temporal irregularity in one of the complex sequences.

Procedure. Participants heard the isochronous cue sequence followed by two complex sequences, which were identical except that one contained a temporal irregularity in the corresponding cued subsequence. Their task was to detect this irregularity by pressing a button indicating whether it was in the first or second complex sequence. They received visual feedback indicating the correctness of their response.

We established the MFS (measured in ERBs) between adjacent subsequences under which listeners could successfully focus on each subsequence separately. "Successful focusing" was defined as the possibility of measuring a threshold with an adaptive staircase procedure (see the Method section of Control Experiment 1 for full details of the 84.1% tracking procedure). We then recorded the temporal irregularity thresholds obtained under these extreme focusing conditions.

The experimental session was as follows. In the single-subsequence condition, detection thresholds were measured for S1 alone (S1 = 700 ms and 234 Hz; in this case, there was no need for a cue sequence). This provided a baseline measure for single-subsequence sequences. In the two-subsequence condition, thresholds were measured for both subsequences (S1 = 700 ms and 234 Hz, S2 = 500 ms and 234 Hz + 1 ERB). If the participant was unable to focus on either subsequence (it was not possible to establish a threshold because the 1 up–4 down staircase did not converge), the difference between the two subsequences was increased by one auditory filter width (ERB), and another attempt was made to establish a threshold. If the participant was still not able to focus, the difference was increased by further steps of 1 ERB until a threshold was established indicating that the participant could focus on both subsequences. We recorded the detection thresholds for both focused subsequences, as well as the MFS (measured in ERBs) required to obtain these thresholds. In the three- and four-subsequence conditions, the procedure was reiterated with three and then four subsequences by adding S3 and S4, respectively, 1 ERB above the last MFS measured. In each case, the frequency differences between each subsequence were increased by steps of 1 ERB until the participant could focus on, and thresholds could be measured for, each subsequence. Once again, we recorded the detection thresholds for both focused subsequences, as well as the MFS (measured in ERBs) required to obtain these thresholds.

As described previously (see control experiments), detection thresholds were measured for temporal irregularities via an adaptive 1 up–4 down procedure (Levitt, 1971). The detection threshold was the degree of temporal irregularity (expressed as a percentage of the target tempo) correctly detected 84.1% of the time. Four successive correct answers resulted in a decrease of 1%; one incorrect answer led to an increase of 1%. Each threshold measurement began with a 10% anisochrony. This value was adopted because pilot studies indicated that thresholds were close to this value, thus increasing accuracy and reducing the length of experimental sessions. The threshold measurement stopped after six reversals (changes in direction), and detection thresholds were calculated by averaging over the last four reversals. On average, 45 trials were required to reach threshold criteria.

For each participant, we obtained 10 thresholds, 1 for each subsequence in each condition (1 for the single-subsequence condition, 2 for the two-subsequence condition, 3 for the three-subsequence condition, and 4 for the four-subsequence condition). We also obtained 12 MFSs (2 in the two-subsequence condition [S1 in relation to S1 = S1/S2 and S2/S1], 4 in the three-subsequence condition [S1/S2, S2/S1, S2/S3, and S3/S2], and 6 in the four-subsequence condition [S1/S2, S2/S1, S2/S3, S3/S2, S3/S4, and S4/S3]). Once thresholds had been obtained for all conditions, the entire procedure was repeated. Listeners completed roughly six 1.5-hr sessions.

Apparatus. The apparatus was the same as in the control experiments.

Participants. Seven listeners took part in this experiment. They all had normal hearing and were paid for their participation. All were undergraduate psychology students at René Descartes University who had not previously participated in psychoacoustic experiments. None of them had a strong musical background (they did not read or play music).

Results

To simplify the presentation of this complex set of data, we have chosen to first present an overview of the results (minimum frequency differences and detection thresholds). Then portions of the data are described to provide answers to specific questions.

Overview of MFS between subsequences. Tables 2, 3, 4, and 5 show all MFS values between adjacent subsequences that participants needed to focus on one subsequence within a complex sequence and, thus, detect the temporal irregularity (i.e., an 84.1% threshold was obtained). The results are expressed as the number of participants needing 1, 2, 3, or 4 ERBs to focus on each subsequence. Tables 2, 3, 4, and 5 show the MFS needed for S1, S2, S3, and S4, respectively, in relation to the possible surrounding subsequences (S2 for S1, S1 and S3 for S2, S2 and S4 for S3, and S3 for S4). In addition, the MFS was calculated for S1 in relation to S2 in three contexts (when there were two, three, or four subsequences in the complex sequence). The results are shown separately for the first and second runs. For instance, when S1 was presented with S2, 6 of the 7 listeners needed only 1 ERB to successfully calculate a threshold. The other listener needed 2 ERBs. This was true for all of the contexts (number of subsequences present in the complex sequence) and for the two runs. Wider MFSs were needed in other conditions; these are discussed later in relation to specific questions.

Overview of detection thresholds. Once the MFS had been established, we recorded the detection thresholds under each experimental condition. Table 6 shows an overview of all mean temporal irregularity detection thresholds (expressed as a percentage of the tempo of the cued subsequence) for each experimental condition (10 conditions created by the four subsequences in different contexts within the complex sequence). A repeated measures ANOVA on the temporal irregularity detection thresholds with subsequence position within the complex sequence (10 levels) and run (2 levels) as variables revealed only a significant effect of subsequence position within the complex sequence, *F*(9,

Table 2
Number of Participants Requiring 1–4 ERBs to Focus on Cued Subsequences: Subsequence 1 (in Relation to Subsequence 2)

		Frequency separation			
Run	Number of subsequences	1 ERB	2 ERBs	3 ERBs	4 ERBs
First	2[a]	6	1	0	0
	3[a]	6	1	0	0
	4[a]	6	1	0	0
Second	2[a]	6	1	0	0
	3[a]	6	1	0	0
	4[a]	6	1	0	0

Note. *N* = 7. The minimum frequency separation was calculated for each subsequence when presented with all other possible adjacent subsequences. Subsequences 1 and 4 were always in an outer position and so were situated only in relation to one other subsequence (above or below), whereas subsequences 2 and 3 could be in either an outer or an inner position and so could be situated in relation to two other subsequences (both above and below). ERB = equivalent rectangular bandwidth.
[a]Cued subsequence was in an outer position.

Table 3
Number of Participants Requiring 1–4 ERBs to Focus on Cued Subsequences: Subsequence 2 (in Relation to Subsequences 1 and 3)

		Subsequence 2–Subsequence 1				Subsequence 2–Subsequence 3			
Run	Number of subsequences	1 ERB	2 ERBs	3 ERBs	4 ERBS	1 ERB	2 ERBs	3 ERBs	4 ERBs
First	2[a]	3	4	0	0				
	3	0	2	5	0	1	0	5	1
	4	0	2	4	1	0	3	4	0
Second	2[a]	6	1	0	0				
	3	3	3	1	0	3	3	1	0
	4	3	3	1	0	1	5	1	0

Note. $N = 7$. The minimum frequency separation was calculated for each subsequence when presented with all other possible adjacent subsequences. Subsequences 1 and 4 were always in an outer position and so were situated only in relation to one other subsequence (above or below), whereas subsequences 2 and 3 could be in either an outer or an inner position and so could be situated in relation to two other subsequences (both above and below). ERB = equivalent rectangular bandwidth.
[a]Cued subsequence was in an outer position.

54) = 23.10, $p < .01$. This effect of position can be analyzed only by breaking it down in relation to the other experimental factors (see following).

Question 1: Single-subsequence and multisubsequence contexts. Is it more difficult to detect temporal irregularities within a subsequence when it is embedded in a multisubsequence context than when it is on its own? If so, this difference would reflect the difficulty of extracting a particular subsequence out of a complex mixture (i.e., a difficulty in selective attending). Figure 3 shows mean temporal irregularity detection thresholds expressed as a percentage of the target tempo in each of the subsequence contexts. When the participants listened to an isolated subsequence, the thresholds were relatively low, at 6.0%. This value was in the same range as those found in our control experiment (5.3%) and in the literature (between 4% and 7%; e.g., Friberg & Sundberg, 1995; Hibi, 1983; Hirsh et al., 1990). It was much lower than those obtained here when the same target subsequence was incorporated into a multisubsequence context (12.5% to 17.6%). A repeated measures ANOVA on the detection thresholds with context (four levels) and run (two levels) as variables revealed a significant effect of context, $F(6, 18) = 70.91$, $p < .01$, but not of run. Planned comparisons revealed a significant difference between the single-subsequence and multisubsequence contexts, $F(1, 6) = 111.02$, $p < .01$, but no significant differences among the three multisubsequence contexts. These results show that when the participants had to direct their attention to one particular subsequence, there was a sizable decrease in temporal performance relative to the single sequence.

Question 2: Relative position. Does the ability to focus on one subsequence depend on the relative position of that subsequence in the complex sequence? Both the MFS and threshold data indicate that it is harder to focus on inner subsequences (S2 in the three-subsequence context, S2 and S3 in the four-subsequence context) than on outer subsequences (S1 and S2 in the two-subsequence context, S1 and S3 in the three-subsequence context, S1 and S4 in the four-subsequence context). First, Table 7 shows that all of

Table 4
Number of Participants Requiring 1–4 ERBs to Focus on Cued Subsequences: Subsequence 3 (in Relation to Subsequences 2 and 4)

		Subsequence 3–Subsequence 2				Subsequence 3–Subsequence 4			
Run	Number of subsequences	1 ERB	2 ERBs	3 ERBs	4 ERBS	1 ERB	2 ERBs	3 ERBs	4 ERBs
First	3[a]	3	4	0	0				
	4	0	3	4	0	1	4	2	0
Second	3[a]	6	1	0	0				
	4	1	5	1	0	2	3	2	0

Note. $N = 7$. The minimum frequency separation was calculated for each subsequence when presented with all other possible adjacent subsequences. Subsequences 1 and 4 were always in an outer position and so were situated only in relation to one other subsequence (above or below), whereas subsequences 2 and 3 could be in either an outer or an inner position and so could be situated in relation to two other subsequences (both above and below). ERB = equivalent rectangular bandwidth.
[a]Cued subsequence was in an outer position.

Table 5
Number of Participants Requiring 1–4 ERBs to Focus on Cued Subsequences: Subsequence 4 (in Relation to Subsequence 3)

Run	Number of subsequences	1 ERB	2 ERBs	3 ERBs	4 ERBs
First	4[a]	4	3	0	0
Second	4[a]	6	1	0	0

Note. *N* = 7. The minimum frequency separation was calculated for each subsequence when presented with all other possible adjacent subsequences. Subsequences 1 and 4 were always in an outer position and so were situated only in relation to one other subsequence (above or below), whereas subsequences 2 and 3 could be in either an outer or an inner position and so could be situated in relation to two other subsequences (both above and below). ERB = equivalent rectangular bandwidth.
[a] Cued subsequence was in an outer position.

the participants needed a small separation (1 or 2 ERBs) to focus on outer subsequences, whereas a wider separation was needed for inner subsequences (half of the participants needed a larger separation of 3 or 4 ERBs). These results for the two-subsequence conditions are similar to those of van Noorden (1975), who found MFSs of between 3 and 5 semitones (1 ERB at 234 Hz corresponds to less than 4 semitones). Similarly, for the same subsequences, thresholds were significantly lower when they were in an outer position (12.6%) than when they were in an inner position (16.6%), an effect confirmed by planned contrasts (subsequences in outer positions vs. the same subsequences in inner positions), $F(1, 6) = 251.03, p < .01$.

Question 3: Number of co-occurring subsequences. Does the number of co-occurring subsequences influence ability to focus on one subsequence? Both the threshold and MFS data indicate that processing was unaffected by the number of co-occurring subsequences. This lack of effect can be seen separately for both the outer and inner subsequences. Table 7 indicates that no differences in MFSs were seen between the different conditions. For the threshold data, Figure 4 indicates no significant difference with the number of distracting subsequences (two-, three-, or four-subsequence contexts). These results suggest that temporal performance is not altered when participants have to focus attention in the presence of an increasing number of distracting subsequences. Even for an inner subsequence, once thresholds could be calculated indicating that listeners could successfully focus on the particular subsequence, detection performance was the same whether there were two or three other subsequences. For outer streams, the selective attention boundary (the point at which streaming is unavoidable) corresponds to less than 1 ERB; for inner streams, it is more than 2 ERBs on average. These limits do not depend on the complexity of the sequence (number of added tones and their temporal relations).

Table 6
Mean Irregularity Thresholds for Subsequences 1–4 in the 1-, 2-, 3-, and 4-Subsequence Contexts

Number of subsequences	Subsequence	*M*	*SD*
1	1	6.0	1.2
2	1	12.8	2.6
	2	11.8	3.7
3	1	12.3	1.8
	2	12.1	2.0
	3	17.5	2.9
4	1	16.0	1.7
	2	13.4	1.7
	3	16.3	1.7
	4	13.3	1.6

Question 4: Practice. Does the ability to focus on a particular stream improve with practice? Diverging results were found with the two sets of data. MFSs were much lower in the second repetition than in the first, showing that focusing can occur with a smaller frequency separation. Given that the entire experiment lasted almost 10 hr, the second repetition was reached after almost three sessions of 1.5 hr. It seems that the selection abilities for a particular subsequence improve with practice. Remember that the thresholds were calculated at the frequency separation at which listeners were just able to successfully focus on the required subsequence. Thresholds were therefore calculated at different MFS levels for the first and second runs but at equal levels of focusing ability. It is therefore not surprising that thresholds did not decrease with practice: A repeated measures ANOVA on thresholds with position (10 levels) and run (2 levels) as variables showed no significant difference for thresholds between the first repetition and the second repetition. So, as predicted by our experimental rationale, the limiting factor to the task is not the ability to detect the temporal irregularity (it is equally easy for all conditions) but, rather, the ability to perceptually organize the complex sequence into streams. Focusing on one particular subsequence is possible only when this organization is successful.

Discussion

Four main results emerged from this experiment. First, it is harder to detect a temporal irregularity in a subsequence when it is embedded within a complex mixture than when it is presented on its own. This finding is in accordance with previous work suggesting that temporal performance in a polyrhythm is generally worse than performance in isochronous sequences (Bharucha & Pryor, 1986), especially below the fission boundary (Jones et al., 1995). Thus, the process of stream segregation implies the use of additional processing effort. We suggest that this involves the perceptual inhibition of nonfocused subsequences; no inhibition is involved in the case of a single subsequence, whereas it is needed when the same subsequence must be separated from other tones in a complex mixture.

Second, for a given subsequence, ease of focusing is unaffected by the number of surrounding subsequences in the complex sequence. Similar results have been obtained with speech stimuli. For instance, Banks and Zender (1984) showed that detection in a shadowed speech message does

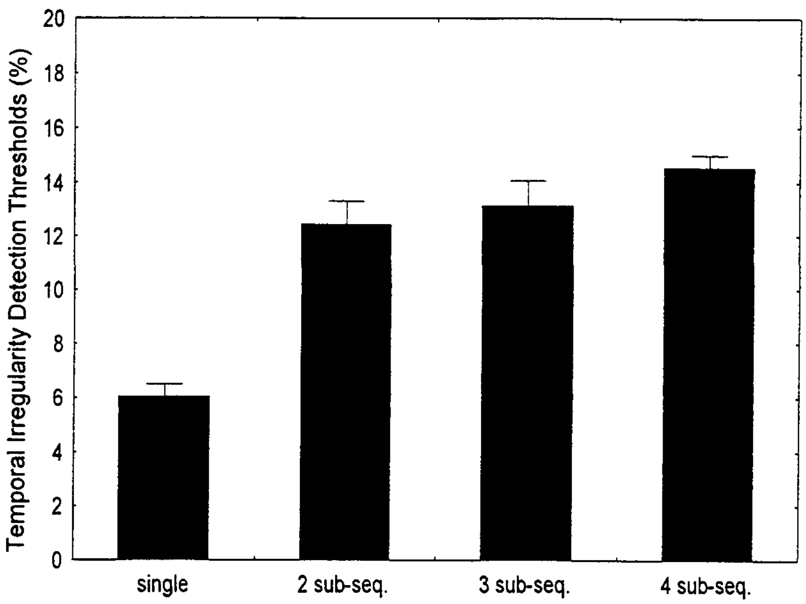

Figure 3. Experiment 1: Mean (*SE*) temporal irregularity detection thresholds (percentage of target subsequence interonset interval) for the four contexts (1, 2, 3, and 4 subsequences). sub-seq. = subsequence.

not decline from one to two distracting messages. Thus, each nonfocused subsequence does not appear to be organized into a single perceptual unit (with as many perceptual units as subsequences); rather, all of the events occurring outside the focused subsequence appear to be grouped into a single perceptual unit making up another complex mixture (slightly less complex than the original mixture because it contains one less subsequence).

Third, it is easier to focus on outer subsequences (highest or lowest) than on inner subsequences. This result is in good agreement with empirical knowledge about the perception of polyphonic music. Listeners, even musicians, have more difficulty detecting entries of inner than outer voices in Bach fugues (Huron, 1989). Also, Huron and Fantini (1989) demonstrated, in an analysis of five-voice figures, that "Bach shows a reluctance to have a new voice enter in an inner position" (p. 43). In a more experimental setting, Palmer and Holleran (1994) showed that pitch changes were easier to detect in outer than in inner musical voices, either in piano or pure-tone sequences. The process of inhibition can account for these findings. Outer subsequences have to be separated only from events in a higher or lower frequency range: Inhibition functions only in one direction. However, inner subsequences have to be separated from events that are both higher and lower in frequency: Inhibition has to function in both directions, considerably increasing the required processing resources.

Fourth, for all experimental conditions, the MFS decreased over the experimental sessions, indicating that the fission boundary is not fixed but, rather, influenced by other factors such as familiarity with the task. The irregularity detection thresholds, however, did not decrease over the experimental sessions, confirming our experimental rationale by which thresholds were measured at the same difficulty level in all conditions: If thresholds had fallen, we would have had to conclude that we were no longer at the fission boundary. The implications of these findings for theories of stream segregation are developed in the General

Table 7
Percentage of Experimental Conditions in Which Listeners Required 1–4 ERBs to Successfully Focus on the Outer and Inner Positions in the 2-, 3-, and 4- Subsequence Contexts

	Outer position				Inner position			
Number of subsequences	1 ERB	2 ERBs	3 ERBs	4 ERBs	1 ERB	2 ERBs	3 ERBs	4 ERBs
2	75	25	0	0				
3	75	25	0	0	25	29	43	3
4	79	21	0	0	14	50	34	2

Note. ERB = equivalent rectangular bandwidth.

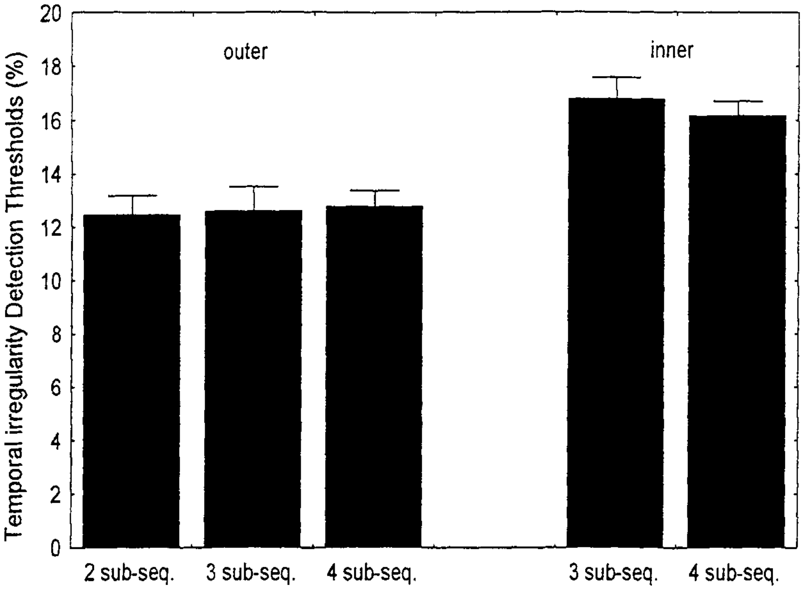

Figure 4. Experiment 1: Mean temporal irregularity detection thresholds (percentage of target subsequence [sub-seq.] tempo) for the outer and inner subsequences as a function of the number of concurrent subsequences.

Discussion section. The differences in focusing on outer and inner subsequences could be explained by a practice effect (the design of the experiment led to more outer than inner conditions), but these differences remained the same for 2 highly trained participants when the tempo–frequency pattern was reversed (data not shown). This problem was addressed in Experiment 2.

Experiment 2: Distinct Zones or Progressive Processes?

In Experiment 1, we investigated perceptual organization under extreme conditions, just above the fission boundary. Each experimental condition began with frequency separations under which listeners were unable to perceptually organize the complex sequence into streams (stream segregation). Only serial integrative listening was possible in instances in which all events were incorporated into a single perceptual unit. Then the frequency separation was increased by small steps until stream segregation was possible (i.e., the fission boundary had been crossed). Thresholds were measured at this point for each listener under each experimental condition. Each threshold was therefore obtained under conditions of equal difficulty. Figure 1C shows the range of experimental conditions used in Experiment 1. In Experiment 2, we investigated how focusing abilities vary as the journey continues upward through the perceptual space toward the temporal coherence boundary (see Figure 2C). A slightly different strategy was adopted from Experiment 1 to reduce the extremely long experimental procedure. Instead of calculating thresholds (with a varying size of temporal irregularity), listeners were required to detect a fixed-size temporal irregularity (15%, well above threshold for the two-subsequence conditions of Experiment 1). The size of the frequency separation was varied between 0.25 and 6 ERBs. We predicted that detection would improve between these two extremes.

Whereas all theories of perceptual organization cited earlier predict an increase in the probability of stream segregation with wider frequency separations, they do not provide clear predictions concerning the shape of this function. At least two functions are possible. If the perceptual space is separated into distinct zones (serial integration below the fission boundary and stream segregation above), there may be an abrupt passage from one to the other, although the point of passage from one state to the other may vary considerably between individuals. In this case, detection should increase rapidly for each listener in each experimental condition and reach a plateau as soon as stream segregation is possible (i.e., as soon as the fission boundary is reached). Alternatively, there may be a progressive increase in the ability to detect the fixed-size temporal irregularity, with the probability of detection increasing with larger frequency separations.

Experiment 2 also allowed us to provide converging support for several of the findings of Experiment 1 under a different paradigm. First, because this experimental task required fewer trials to obtain stable results, we were able to test more situations. Therefore, our goal was to verify that the results obtained with the particular frequency–tempo combination used in Experiment 1 (the fastest subsequences were always the highest: S1–700 ms, S2–500 ms, S3–400 ms, and S4–300 ms) generalize to other frequency–tempo combinations. If this is the case, we can conclude that the underlying processes function irrespective of the context. In

Experiment 1, because it appeared that structural information did not influence detection performance, we assumed that nonfocused subsequences were not processed into separate streams. According to this hypothesis, the reversal of the temporal context should not influence stream segregation in the present experiment.

Second, we wished to confirm the relative ease of focusing on outer voices as compared with inner voices. If this is the case, detection performance should increase faster (for smaller frequency separations) for the former than for the latter.

Third, we investigated the role of musical expertise in the ability to organize complex sequences into streams. As discussed earlier, we expected increased stream segregation skills in musicians who have developed expert listening skills through years of specific training. Indeed, Jones et al. (1995) found that musically trained listeners performed better than musically naive listeners in a selective listening task. Experiment 1 indicated that MFS decreased with practice. Thus, frequency constraints on selective processes may vary with increasing skill level. This difference may be due, in part, to shifted perceptual boundaries. We predicted a shift of the fission boundary toward smaller frequency separations for the expert listeners relative to nonmusicians.

Method

Stimuli. Sequences were composed of three subsequences of pure tones. The frequency could be low, middle, or high, and the tempo could have an IOI of 700 ms, 500 ms, or 300 ms. The target subsequence that contained the temporal irregularity always had an IOI of 500 ms, but the frequency varied from trial to trial, so the target could be the highest, middle, or lowest subsequence. For each target subsequence, we examined two contexts. For example, in the first context for the middle target, the highest frequency subsequence was the fastest (300 ms), and the lowest frequency subsequence was the slowest (700 ms). The second context was composed of the other possible pattern. Each combination was repeated five times in a block (30 trials) and run in a counterbalanced order. Each pair of complex sequences was preceded by a single subsequence (cue sequence) that had the same tempo and frequency as the target subsequence.

In a block of trials, the subsequences all had the same frequency separation. This separation corresponded to 0.25, 1, 2, 4, or 6 ERBs above or below 659 Hz (0.25 ERBs, 635 and 683 Hz; 1 ERB, 568 and 760 Hz; 2 ERBs, 486 and 872 Hz; 4 ERBs, 346 and 1137 Hz; and 6 ERBs, 234 and 1465 Hz). The size of the frequency separation (five levels) was counterbalanced between blocks across the sessions. The temporal irregularity could occur in one of three positions: near the beginning, middle, or end of the sequence (as in Experiment 1). The size of the temporal irregularity was fixed at a value of 15% of the target tempo (500 ms) according to a preliminary test. This value was well above the thresholds found for the outer subsequences (12%, at an 84.1% detection level) in Experiment 1. Nonfocused subsequences never contained temporal irregularities. Tones had the same duration and level as in Experiment 1. The total duration of one sequence was 3,650 ms. In all, listeners completed 300 trials (5 levels × 3 positions × 2 contexts × 10 repetitions).

Participants. Twenty-two listeners took part in this experiment. They all had normal hearing and were paid for their participation. They had not previously taken part in psychoacoustic experiments. Eight had a musical background (varying from 7 to 15 years of musical training) and played a musical instrument at least 2 hr weekly. These listeners were included in the analyses as musicians. In the nonmusician group, participants did not read music or play a musical instrument.

Apparatus. The apparatus was the same as in the previous experiments.

Results

A mixed ANOVA on the percentage of correct irregularity detections was performed. The between-subjects variable was expertise (musicians vs. nonmusicians), and the within-subjects variables were frequency separation (five levels), position of the target (low vs. middle vs. high), and frequency–tempo context (two levels). Each of these effects is analyzed in turn.

Figure 5 shows the detection rate as a function of the frequency separation for the three positions (high, middle, and low) and for the nonmusicians and musicians. In general, the predicted increase in detection performance with frequency separation was observed: significant main effect of frequency separation, $F(4, 80) = 75.58, p < .001$. At one extreme, detection was at chance level (between 40% and 60% correct = 50% ± mean standard deviation) when all of the tones were within the same critical band (0.25 ERBs). At the other extreme, detection was optimal (above 90% correct) with a frequency separation of about 4 ERBs or more. Between these two extremes, detection increased gradually with increased frequency separation.

In accordance with the results of Experiment 1, it was easier to detect the temporal irregularity for the outer (highest [77%] and lowest [83%]) than for the inner (middle [68%]) subsequences: significant main effect of position, $F(2, 40) = 40.96, p < .001$. The general pattern of increase in detection with increased frequency separation was observed for all positions. The significant interaction between frequency separation and position, $F(8, 160) = 3.73, p < .001$, arose from the fact that the three curves started together at 0.25 ERBs, diverged at intermediate values, and then reconverged at the largest values.

Of particular importance is the absence of a significant main effect of frequency–tempo context, $F(1, 20) = 2.57, p = .13$, or its interaction with other variables. This finding confirms that the ability to organize complex sequences into streams is unaffected by the physical characteristics of surrounding subsequences (at least for the tempo combinations used in these experiments).

Overall, musicians performed better than nonmusicians: main effect of expertise, $F(1, 20) = 8.57, p < .009$. There were no significant interactions with other variables. Figure 5 shows how all curves were shifted to the left for the musicians relative to the nonmusicians.

Table 8 presents the interpolated frequency separation required to pass above the mean chance level (60% correct: 50% ± 1 standard deviation, perhaps an indication of the fission boundary), at a 75% threshold level and at the constructed optimum detection level (90% correct; see Method section), for the three positions and two groups of listeners. The interpolations were based on the function $P =$

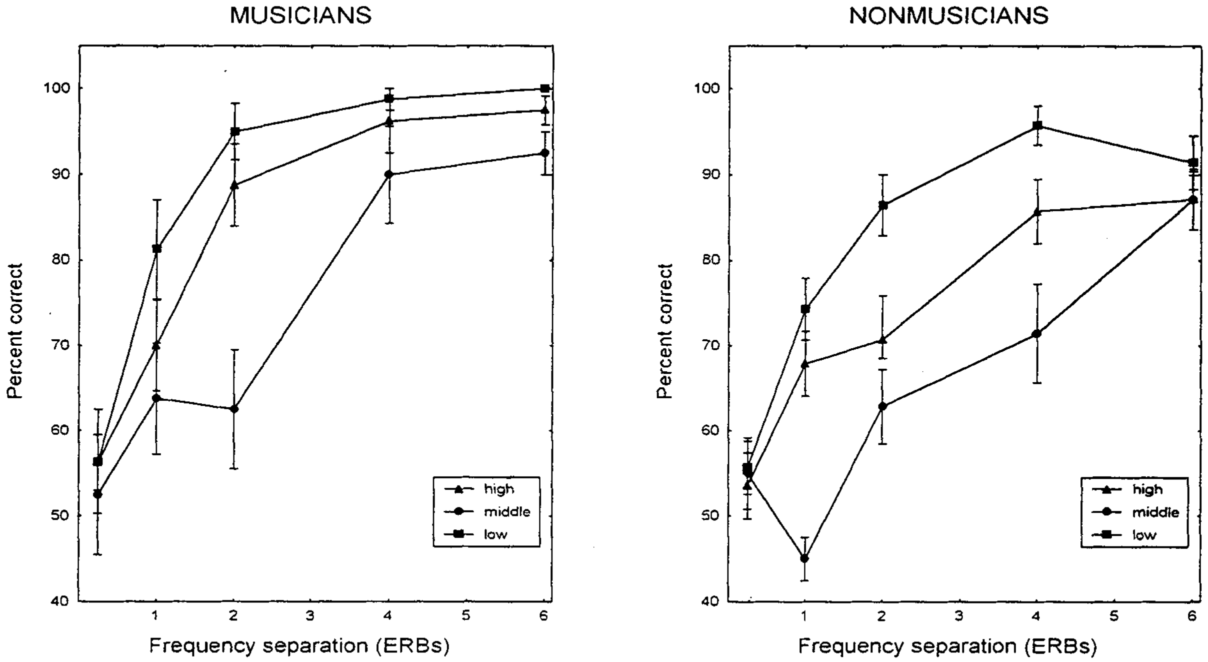

Figure 5. Experiment 2: Percentages of correct detections (*SEM*) as a function of frequency separation for the highest, middle, and lowest subsequences among musicians (n = 8) and nonmusicians (n = 14). ERBs = equivalent rectangular bandwidths.

$1 - 2[-(I/\alpha)\beta]$, where α is threshold and β is slope. These data illustrate the rightward shift of the probability of detection for nonmusicians relative to musicians and for middle subsequences relative to outer subsequences.

The smooth curves in Figure 5 showing a gradual increase in performance with frequency separation were obtained by averaging over the listeners in each group. It could be that each listener's curve shows an abrupt passage from one state to the other (integrative listening–streaming) resulting in a step function but that, combined, they give the gradual curves due to the distribution of individual transition thresholds. Examination of individual curves for each condition provides an indication that this is not the case. An example is provided in Figure 6: Individual curves for the lowest frequency condition are presented according to the number of frequency separations over which the increase occurred.

Table 8

Interpolated Frequency Separations Observed for the 60% (Chance), 75% (Threshold), and 90% (Maximum) Levels for the Two Groups of Listeners and the Three Positions

Group and position	60%	75%	90%
Musicians			
High	0.3	1.1	2.6
Middle	1.0	2.6	5.0
Low	0.2	0.7	1.6
Nonmusicians			
High	0.5	2.3	6.7
Middle	2.0	4.0	7.6
Low	0.2	1.0	3.5

Of the 22 listeners, 1 showed an erratic function, 3 showed a one-step increase, 9 showed a two-step increase, 7 showed a three-step increase, and 2 showed a four-step increase. Therefore, 18 listeners showed an increase over two or more classes, providing support for the hypothesis of a progressive function. Only 3 listeners showed a one-step function consistent with the "abrupt change" hypothesis; because all 3 fell between 0.25 ERB and 1 ERB, however, a more fine-grained scale toward the smaller frequency separations would probably reveal a more progressive function. These data therefore suggest a progressive increase in the probability of stream segregation rather than an abrupt change from one state to the next.

Finally, Table 8 may shed light on factors that limit perceptual organization. If chance level is taken as an indication of the passage above the fission boundary, note that, with one exception (nonmusicians, middle subsequence), all extrapolated boundaries occurred at roughly the same point (at or just below 1 ERB). Thus, the fission boundary appears to be limited by early processing (referred to as "primitive analysis" by Bregman, 1990) and is only slightly affected by more top-down processes such as attention and skill level. However, the 75% and 90% limits were much more variable, suggesting that once the fission boundary is surpassed, the probability of stream segregation may be more influenced by these top-down processes. Planned contrasts testing for differences between positions and groups indicated no differences for the smallest frequency separation but significant differences for all of the other separations. Indeed, Fine and Moore (1993) showed that musicians are better at hearing one partial tone out of a

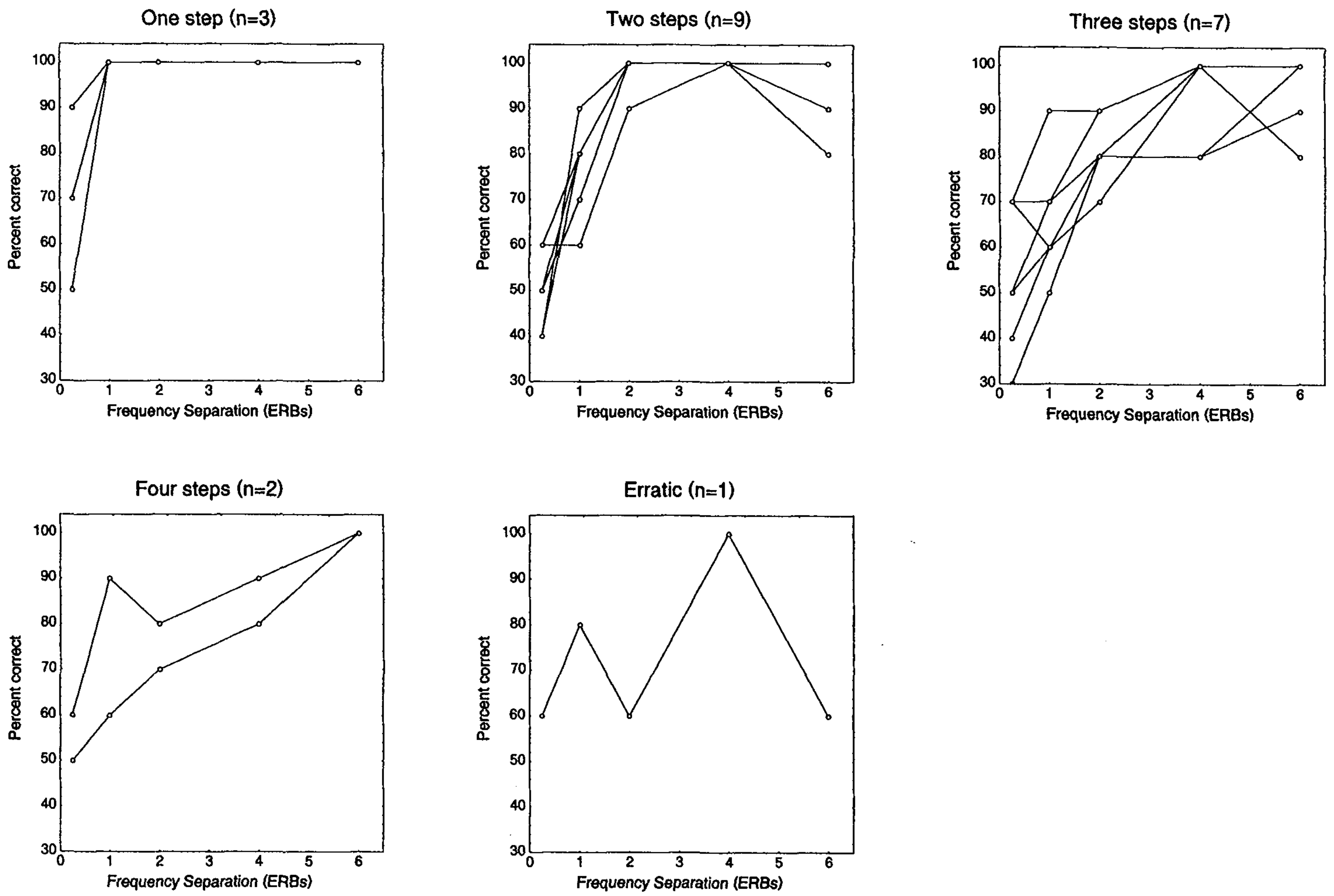

Figure 6. Experiment 2: Individual data for the lowest subsequence. Listeners are organized in relation to the number of frequency steps over which the increase in detection performance occurred. ERBs = equivalent rectangular bandwidths.

complex tone but that the size of their auditory filters does not differ significantly from that of nonmusicians, although Spiegel and Watson (1984) showed better frequency discrimination sensitivity by professional musicians.

Discussion

This experiment produced four main results. First, we have created experimental conditions under which listeners travel through are tested at several conditions of stimulus parameters in the stream segregation perceptual space. At one extreme, stream segregation is impossible, with listeners perceiving the complex sequence in an integrated fashion. They are therefore unable to focus on a particular subsequence or detect the temporal irregularity within that subsequence. As frequency separation increases, listeners pass over the fission boundary and start being able to organize the complex sequence into two perceptual units (the focused stream and the unfocused mixture). As stream segregation becomes possible, listeners become increasingly able to focus on one subsequence and detect the temporal irregularity within it. Both the group data and the individual data indicate that this probability increases progressively as a function of frequency separation; there is no abrupt passage from one state to the other as suggested from Experiment 1. In addition, it seems likely that the temporal coherence boundary was reached, because detection was almost perfect above a frequency separation of about 4 ERBs (depending on the experimental conditions).

Second, detection was better (the curves shifted to the left) for the outer subsequences than for the inner subsequences. This confirms the results of Experiment 1 with a different paradigm.

Third, the ability to focus on a particular stream is unaffected by the surrounding frequency–tempo combination, suggesting that the underlying processes function irrespective of context. This finding supports the hypothesis that nonfocused subsequences are not processed in separate perceptual units but, rather, represent an undifferentiated background. The results of Experiment 1 can therefore be generalized to other frequency–tempo combinations.

Fourth, musicians were better able than nonmusicians to organize the complex sequences into streams: All of the experimental curves were shifted toward the left. Whereas the fission boundary appears to be relatively fixed, the probability of stream segregation (and perhaps the temporal coherence boundary) above this limit appears to be less rigid and more susceptible to influence by top-down processes such as attention, which may in turn be influenced by skill level. Therefore, our prediction that the fission boundary is shifted toward smaller frequency separations has not been confirmed; only the probability of stream segregation above this boundary is increased.

General Discussion

In this article, we have examined how listeners perceptually organize complex sequences of two or more concurrent subsequences. We have focused on the way in which the frequency separation between adjacent subsequences influences stream segregation. The pattern of results allowed us to deduce certain underlying processes. A companion article focusing on the role of temporal organization in these processes is currently in preparation.

Generalizing the Principles of Stream Segregation

Traditionally, studies have concentrated on the perceptual space of two potential streams with three zones (below the fission boundary, above the temporal coherence boundary, and between the two). Results from these studies indicate that, to a certain extent, these concepts can be applied to more complex situations. Further support has been provided for two principles behind stream segregation.

Principle 1: As for two-subsequence mixtures, stream segregation occurs only if adjacent subsequences are sufficiently separated in frequency. Our data indicate a required separation of about 3 semitones (1 ERB), consistent with those in the literature that range from 3–5 semitones. This is consistent with Bregman's suggestion that this limit is related to physiological limits affecting the frequency resolution of the auditory system (auditory filters). Such limits have been successfully incorporated in computational modeling of auditory stream analysis (Beauvois & Meddis, 1996).

Principle 2: Focusing on a subsequence within a complex mixture requires attentional effort. Irregularity detection thresholds were higher in complex mixtures than in single sequences. Organizing the mixture into focused and unfocused streams requires additional cognitive (attentional?) resources.

We add four new principles concerning the generalizability of previous knowledge obtained with two subsequences to more complex contexts.

Principle 3: Stream segregation is not an all-or-none phenomenon with an abrupt passage from a state of serial integration into a state of stream segregation. Instead, there is a gradual increase in the probability of a segregated percept occurring as the frequency separation increases. This is probably also the case above the temporal coherence boundary. This probability of segregation does not imply that, in the ambiguous region of the perceptual space, listeners experience a mixed perception (segregation plus integration): One cannot hear 1.57 streams! Thus, there is a probability gradient between the two processes (stream segregation and serial integration) that reflects the degree of attentional effort required to obtain one or the other percept. This implies a single process with two possible perceptual results, the probability of occurrence of which is determined in large part by two properties of the stimulus configuration (frequency and temporal proximity; Jones & Yee, 1993).

Principle 4: The probability of stream segregation depends on the position of the to-be-focused subsequence within the complex mixture. Two-subsequence mixtures contain only outer subsequences that have to be separated from other subsequences in one direction (higher or lower frequency). We suggest that a process of perceptual inhibition is functioning that requires attentional effort. The focusing on inner subsequences in more complex mixtures involves their separation from both higher and lower frequency events. The fact that it is more difficult to focus on inner than outer subsequences suggests that inhibition is necessary in both directions, a situation requiring additional attentional effort.

Principle 5: The probability of successfully focusing on a particular subsequence is unaffected by the physical characteristics of surrounding events (excluding the frequency separation between subsequences); similar probabilities were observed with different frequency–tempo combinations. This result sheds light on the issue of the depth of processing of nonfocused subsequences. It suggests that nonfocused subsequences are not processed in as much depth as focused ones, because they are not organized into separate streams (but see Alain & Woods, 1993, 1994, for an alternative interpretation). Further studies are currently addressing this issue.

Principle 6: Whereas the fission boundary may be fixed by physiological constraints, the probability of stream segregation is probably more susceptible to influence by top-down processes such as attentional set and learning (as suggested by Bregman, 1990). Indeed, the fission boundary did not vary much between musician and nonmusician listeners, but the probability of stream segregation above the fission boundary did: Musicians were more likely to be able to focus on a particular subsequence for a smaller frequency separation.

Toward a Model of Stream Segregation in Complex Sequences

We propose the first version of a model of the perceptual organization of complex sequences composed of pure tones. We suggest that when listeners focus their attention on an outer subsequence, they divide the auditory environment into two spaces: a first perceptual unit corresponding to the focused subsequence and a second perceptual unit composed of all other tones. The tones in the unfocused perceptual unit are not organized into separate streams but coded in a global fashion. The situation is more complex when listeners focus on inner subsequences because they divide their environment into three spaces: The first is composed of the focused subsequence, the second is composed of all tones lower in frequency, and the third is composed of all tones higher in frequency.

However, this perceptual organization is not fixed or rigid, relying only on frequency separation as may be inferred from the preceding discussion. Other factors, both bottom-up and top-down, may influence the perceived salience of particular subsequences, thus changing the perceptual organization of the complex mixture and leading to different perceptual units. First, physical tone characteristics (such as intensity, tempo, and timbre) influence the relative salience of subsequences (Handel, 1984; Spiegel & Watson, 1981). For instance, Botte et al. (1997) demonstrated that an increase in level of about 15 dB changed a previously unfocused subsequence into a focused stream (listeners were able to detect temporal irregularities within that subsequence). Second, top-down schemes, particularly those related to attentional selection, can change perceptual organization (Dowling, 1973; Dowling et al., 1987). For instance, focusing attention on a particular aspect of a complex mixture can increase the salience of a normally unsalient subsequence sufficiently for it to become a single perceptual unit.

We have discussed some perceptual mechanisms that could be involved in this type of organization. First, inhibition of nonfocused sound events may occur through a process of perceptual attentuation discussed in detail elsewhere (Botte et al., 1997). We have chosen to describe the active processes that result in an enhanced perception of the focused stream in relation to the unfocused subsequences in terms of the inhibition (increased thresholds) of nonfocused subsequence, but we acknowledge the alternative interpretation of facilitation (reduced thresholds) of the focused stream. Choosing to focus on one aspect of the sequence is the other side of the coin of choosing to ignore other aspects of the sequence. This is a long-standing and still unresolved debate in the field of psychoacoustics. The present article does not provide any new insight into this particular debate.

Second, we suggest that attentional processes are involved at an early stage in the perceptual organization of complex auditory sequences. These attentional processes highlight the important, relevant information, leading to enhanced processing of the "selected" events. One outcome of this process is that the "nonselected" events are processed to a far lesser degree. We have gone as far as to suggest that they are not perceptually organized into separate perceptual units (or streams) but, rather, processed in a single, nondifferentiated mixture composed of all nonfocused events. This view is contrary to the prevailing view proposed by Bregman that automatic streaming processes result in the creation of all possible streams at an early stage in processing and that attention allows the listener to select one of these streams. Our data do not allow us to arbitrate definitively between these two theoretical positions. Recent evoked potential data do, however, provide evidence in favor of the hypothesis that attention intervenes at a very early stage of processing, directly influencing perceptual organization of the complex sequence (Alain & Woods, 1993, 1994; Alain, Achim, & Richer, 1993; Sussmann, Ritter, & Vaughan, 1998, in press). We hope that the creation of new experimental paradigms will provide new insights into this question in the near future.

Third, we follow Bregman in the proposition that stream segregation based on frequency separation is determined by physiological constraints of the cochlea. This low-level physiological mechanism determines the lower limits of what can possibly be perceived as a stream. Attentional selection processes cannot go beyond this lower limit. The temporal structure of sequences clearly influences stream segregation but probably not at such a peripheral level.

In this regard, it would be of particular interest to test this model of perceptual organization with sounds of different spectral composition. Palmer and Holleran (1994) showed that preferred focusing on the highest voice slightly decreases when pure tones, rather than piano sounds are used.

According to Van Noorden (1975), the temporal coherence boundary might have a sharper slope with complex tones, but recent investigations (Bey & McAdams, 1997) have shown that the perceptual limits of sequential organization might be the same for either pure or complex sounds.

Our findings with relatively artificial complex sequences can be compared with those obtained by Huron (1989) with real music in which tones differ along many different dimensions. He concluded that the auditory system seems to follow a "one, two, three, or many" rule; that is, it may be impossible to process more than four concomitant streams. Our results suggest that, when subsequences differ only by two characteristics (tempo and frequency), this limit may not be more than three.

References

Alain, C., Achim, A., & Richer, F. (1993). Perceptual context and the selective attention effect on auditory event-related brain potentials. *Psychophysiology, 30,* 572–580.

Alain, C., & Woods, D. L. (1993). Distractor clustering enhances detection speed and accuracy during selective listening. *Perception & Psychophysics, 54,* 509–514.

Alain, C., & Woods, D. L. (1994). Signal clustering modulates auditory cortical activity in humans. *Perception & Psychophysics, 56,* 501–516.

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

Banks, W. P., & Zender, J. P. (1984). A test of time sharing in auditory attention. *Bulletin of the Psychonomic Society, 22,* 541–544.

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

Bey, C., & McAdams, S. (1997). Implication des processus descendants dans la formation des flux auditifs [Implication of top-down processes in auditory stream formation]. *Actes des Journées Internationales d'Orsay de Sciences Cognitives,* 29–32.

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

Botte, M.-C., Drake, C., Brochard, R., & McAdams, S. (1997). Preliminary measures of the focusing of attention on auditory streams. *Perception & Psychophysics, 59,* 419–425.

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 in rapid sequences of tones. *Journal of Experimental Psychology, 89,* 244–249.

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

Davidson, B., Power, R. P., & Michie, P. T. (1987). The effects of familiarity and previous training on perception of an ambiguous musical figure. *Perception & Psychophysics, 41,* 601–608.

Dowling, W. J. (1973). The perception of interleaved melodies. *Cognitive Psychology, 5,* 322–337.

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

Drake, C., & Botte, M. C. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Perception & Psychophysics, 54,* 277–286.

Fine, P. A., & Moore, B. C. J. (1993). Frequency analysis and musical ability. *Music Perception, 11,* 39–54.

Fraisse, P. (1967). Le seuil différentiel de durée dans une suite régulière d'intervalles [Temporal thresholds for regular sequences]. *L'Année Psychologique, 1,* 43–49.

Friberg, A., & Sundberg, J. (1995). Time discrimination in a monotonic, isochronous sequence. *Journal of the Acoustical Society of America, 98,* 2524–2531.

Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of ERB shape from notched-noise data. *Hearing Research, 47,* 103–138.

Handel, S. (1984). Using polyrhythms to study rhythm. *Music Perception, 1,* 465–484.

Hibi, S. (1983). Rhythm perception in repetitive sound sequences. *Journal of the Acoustical Society of Japan, 4,* 83–95.

Hirsh, I. J., Monahan, C. B., Grant, K. W., & Singh, P. G. (1990). Studies in auditory timing I: Simple patterns. *Perception & Psychophysics, 47,* 215–226.

Huron, D. (1989). Voice denumerability in polyphonic music of homogeneous timbres. *Music Perception, 6,* 361–382.

Huron, D., & Fantini, D. (1989). The avoidance of inner-voice entries: Perceptual evidence and musical practice. *Music Perception, 7,* 43–48.

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review, 83,* 323–355.

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review, 96,* 459–491.

Jones, M. R., Jagacinski, R. J., Yee, W., Floyd, R. L., & Klapp, S. T. (1995). Test of attentional flexibility in polyrhythmic patterns. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 293–307.

Jones, M. R., Maser, D. J., & Kidd, G. R. (1978). Rate and structure in memory for auditory patterns. *Memory & Cognition, 6,* 246–258.

Jones, M. R., & McCallum, R. (1987). An application of principal directions in scaling to auditory pattern perception. In F. Young & R. Hamer (Eds.), *Multidimensional scaling: Theory and applications* (pp. 259–278). Hillsdale, NJ: Erlbaum.

Jones, M. R., & Yee, W. (1993). Attending to auditory events: The role of temporal organization. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 69–112). Oxford, England: Clarendon Press.

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

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

Palmer, C., & Holleran, S. (1994). Harmonic, melodic and frequency height influences in the perception of multivoiced music. *Perception & Psychophysics, 56,* 301–312.

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

Scharf, B., & Houstma, A. J. M. (1986). Audition II. Loudness, pitch, localization, aural distortion, pathology. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 1; pp. 15.1–15.60). New York: Wiley.

Smith, J., Hausfeld, S., Power, R. P., & Gorta, A. (1982). Ambiguous musical figures and auditory streaming. *Perception & Psychophysics, 32,* 454–464.

Spiegel, M. F., & Watson, C. S. (1981). Factors in the discrimination of auditory patterns III: Frequency discrimination with well-learned patterns. *Journal of the Acoustical Society of America, 69,* 223–230.

Spiegel, M. F., & Watson, C. S. (1984). Performance on frequency-discrimination tasks by musicians and nonmusicians. *Journal of the Acoustical Society of America, 76,* 1690–1695.

Sussmann, E., Ritter, W., & Vaughan, H. G. (1998). Attention affects the organization of auditory input associated with the mismatch negativity system. *Brain Research, 789,* 130–138.

Sussmann, E., Ritter, W., & Vaughan, H. G. (in press). Investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology.*

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 October 13, 1997
Revision received July 7, 1998
Accepted December 8, 1998 ■

New Editors Appointed, 2001–2006

The Publications and Communications Board of the American Psychological Association announces the appointment of seven new editors for 6-year terms beginning in 2001. As of January 1, 2000, manuscripts should be directed as follows:

- For the **Journal of Abnormal Psychology,** submit manuscripts to Timothy B. Baker, PhD, Department of Psychology and CTRI, 7255 Medical Sciences Center, 1300 University Avenue, University of Wisconsin Medical School, Madison, WI 53706.
- For the **Journal of Comparative Psychology,** submit manuscripts to Meredith West, PhD, Department of Psychology, 1101 E. 10th Street, Indiana University, Bloomington, IN 47405-7007.
- For the **Journal of Experimental Psychology: Learning, Memory, and Cognition,** submit manuscripts to Thomas O. Nelson, PhD, Psychology Department, University of Maryland, College Park, MD 20742-4411.
- For the **Journal of Personality and Social Psychology: Attitudes and Social Cognition** section, submit manuscripts to Patricia Devine, PhD, Department of Psychology, University of Wisconsin—Madison, 1202 West Johnson Street, Madison, WI 53706-1696.
- For **Professional Psychology: Research and Practice,** submit manuscripts to Mary Beth Kenkel, PhD, California School of Professional Psychology—Fresno, 5130 East Clinton Way, Fresno, CA 93727.
- For **Psychological Review,** submit manuscripts to Walter Mischel, PhD, Department of Psychology, 406 Schermerhorn Hall, Columbia University, New York, NY 10027.
- For **Psychology, Public Policy, and Law,** submit manuscripts to Jane Goodman-Delahunty, JD, PhD, 2407 Calle Madiera, San Clemente, CA 92672.

Manuscript submission patterns make the precise date of completion of the 2000 volumes uncertain. Current editors, Milton E. Strauss, PhD; Charles T. Snowdon, PhD; James H. Neely, PhD; Arie Kruglanski, PhD; Patrick H. DeLeon, PhD, JD; Robert A. Bjork, PhD; and Bruce D. Sales, JD, PhD, respectively, will receive and consider manuscripts through December 31, 1999. Should 2000 volumes be completed before that date, manuscripts will be redirected to the new editors for consideration in 2001 volumes.