Journal of Experimental Psychology:
Human Perception and Performance
2003, Vol. 29, No. 2, 267–279

0096-1523/03/$12.00 DOI: 10.1037/0096-1523.29.2.267

Postrecognition of Interleaved Melodies as an Indirect Measure of Auditory Stream Formation

Caroline Bey and Stephen McAdams
Université René Descartes and Institut de Recherche et Coordination Acoustique/Musique–
Centre National de la Recherche Scientifique

Primitive processes involved in auditory stream formation are measured with an indirect, objective method. A target melody interleaved with a distractor sequence is followed by a probe melody that was identical to the target or differed by 2 notes. Listeners decided whether the probe melody was present or not in the composite sequence. Interleaved melody recognition is not possible when distractor sequences have the same mean frequency and maximum contour crossover with target melodies. Performance increases with mean frequency separation and timbral dissimilarity and is unaffected by the duration of the silent interval between composite sequence and probe melody. The relation between this indirect task measuring the interleaved melody recognition boundary and direct judgments measuring the fission boundary is discussed.

The everyday auditory environment consists of multiple, simultaneously active sources with overlapping temporal and spectral acoustic properties. Despite the seemingly chaotic composite signal impinging on people's ears, the resulting perception is of an orderly "auditory scene" that is organized according to sources and auditory events. Individuals are thus able to follow speech in a noisy environment (Cherry, 1953) and to isolate melodic voices in polyphonic music. According to Bregman (1990, 1993), the analysis of an auditory scene is preattentive in part, and he referred to this component as "primitive analysis." The complex signal arising from various acoustic sources that reaches people's ears is decomposed in a preattentive way into independent perceptual entities, called "auditory streams," which generally correspond to the different sources of the environment. The auditory system uses regularly occurring acoustic cues to build streams, like the harmonicity of many relevant sounds of the environment, the asynchrony of independent sources, and the smooth change over time of sound properties (e.g., fundamental frequency, spectral and temporal envelope) coming from the same source. One main argument supporting the hypothesis that auditory stream formation involves preattentive processes is the fact that segregation may occur even when sounds are not attended to (Sussman, Ritter, & Vaughan, 1999) and even against listeners' intentions (van Noorden, 1975). Moreover the ability to build streams is present very early in life (McAdams & Bertoncini, 1997) and shared by other species like birds (MacDougall-Shackleton, Hulse, Gentner, & White, 1998), suggesting that it is innate and adaptive. It should be noted nonetheless that evidence has been found for the involvement of attentional processes in stream formation (Alain & Woods, 1994; Carlyon, Cusack, Foxton, & Robertson, 2001; Sussman, Ritter, & Vaughan, 1998; van Noorden, 1975). In this article, we address the problem of measuring the primitive processes involved in auditory stream formation.

The majority of studies on auditory stream formation processes have examined the switching between two percepts when a sequence organized into one stream splits into two streams or the reverse. This "switching" phenomenon, measured using threshold techniques, has been studied with fairly simple cyclical sound sequences, in which the factors that induce a change in perceptual analysis are manipulated. Isochronous cyclical sequences composed of two alternating sounds—A and B, which have different characteristics (frequency, spectral content, etc.)—have mainly been used. They were either of the ABAB type (Miller & Heise, 1950; van Noorden, 1975, Experiment 2) or of the ABA—ABA— type, where — indicates a silence (Rogers & Bregman, 1998; Singh & Bregman, 1997; van Noorden, 1975, Experiment 1; Vliegen & Oxenham, 1999). The ABAB sequence is heard either as a trill alternating between two notes (ABAB) when one stream is formed (referred to as *integration*) or as two separate streams (A—A—A— and B—B—B—), each with a single note repeated at half the tempo of the integrated trill percept (Figure 1a). The ABA—ABA— sequence produces a different rhythm, depending on whether the tones are organized into one or two streams: a galloping rhythm (ABA—ABA—) or two isochronous, single-note streams (A—A—A— and B————B————), one occurring at twice the tempo of the other (Figure 1b). Note that listeners are often asked to focus on the perceived rhythm to decide whether one or two streams are being heard. In both stimuli, the segregated percept is referred to as *fission.*

Sequences composed of repeating cycles of more than two different sounds have also been used (Bregman & Campbell, 1971; Bregman, Liao, & Levitan, 1990; McAdams & Bregman, 1979; Singh, 1987). The use of cyclical sequences has several advantages (Bregman, 1990, p. 53): (a) long sequences enhance the fission

Caroline Bey and Stephen McAdams, Laboratoire de Psychologie Expérimentale, Centre National de la Recherche Scientifique (CNRS), Université René Descartes, Paris, France, and Institut de Recherche et Coordination Acoustique/Musique–CNRS, Paris, France.

Preliminary results leading to this article were presented at the Fourth Congrès Français d'Acoustique, Marseille, France, April 1997.

Correspondence concerning this article should be addressed to Stephen McAdams, IRCAM–CNRS, 1 place Igor Stravinsky, F-75004 Paris, France. E-mail: smc@ircam.fr

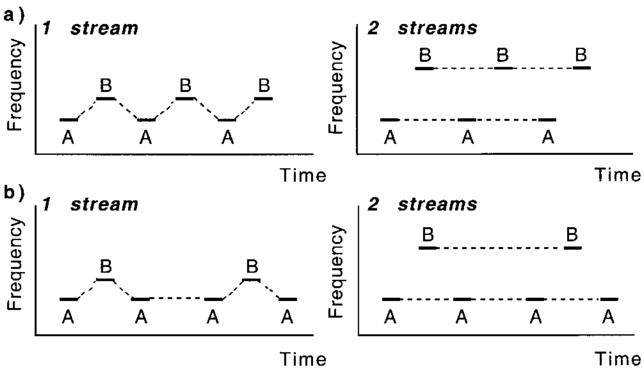

Figure 1. Schematic representation of the two types of cyclical sequences frequently used to study auditory stream formation: (a) ABAB sequence and (b) ABA—ABA— sequence. — indicates a silence.

phenomenon, which results from a cumulative process (Bregman, 1978a; Rogers & Bregman, 1998); (b) the structure remains constant, allowing a simple description; and (c) this type of pattern minimizes influences of the beginning and end of the sequence on the response, which is important when the task involves reporting on the order of the sounds, for example (Bregman & Campbell, 1971). However, these stimuli have the major disadvantage of being highly predictable. The repetition of the same sound pattern allows a listener to build a representation of the sequence, which may then intervene as knowledge in top-down fashion in the building of streams. This schema-based analysis distinguished by Bregman (1990, chap. 4; 1993) is a selection process. The analysis does not consist of a partition leading to the building of streams but in a matching process between the activated knowledge stored in memory and the sensory representation of the incoming signal. Moreover, the method, which consists of directly asking listeners what they perceive (one or two streams), requires many observations to obtain coherent results because of its subjectivity and thus its variability.

Indirect methods that consist of recognizing one target in a complex pattern have also been used to measure stream formation. Dowling (1973; Dowling, Lung, & Herrbold, 1987) proposed an interleaved melody recognition task, subsequently used by several authors (Gregory, 1994; Hartmann & Johnson, 1991; Iverson, 1995; Vliegen & Oxenham, 1999, Experiment 2), in which listeners had to judge if a familiar tune or an unfamiliar melody presented in a first sequence was present or not in a second sequence composed by a melody interleaved with distractor tones. Researchers also used other target properties like rhythm (Singh & Bregman, 1997; Vliegen, Moore, & Oxenham, 1999) or the order of the sounds (Bregman & Campbell, 1971). These indirect methods measure the representation of the target stream properties rather than the percept change (the switch between one to two streams or the reverse). The number of streams built by the auditory system is thus inferred on the basis of performance in recognition of target patterns. These methods provide an objective measure of auditory stream formation because the response given by the listener can be compared with the actual stimulus. The number of observations needed is also less than in direct methods, which represents an advantage for clinical testing in audiology and neurology. However, trying to recognize a target pattern heard immediately before or stored in long-term memory involves top-down processes.

Studies using direct methods to access the auditory scene representation have established that listeners are able to segregate a sequence of two pure tones alternating between different frequencies into two streams when the frequency difference is around 15% (Miller & Heise, 1950) or 2–3 semitones (ST) for a frequency of 1000 Hz (fission boundary; Miller & Heise, 1950; van Noorden, 1975, Experiment 1). However, the results of studies using indirect methods have shown that a mean fundamental frequency difference of at least one octave (12 ST) is necessary to recognize a melody interleaved with distractor sounds (Dowling, 1973, Experiment 1; Hartmann & Johnson, 1991; Vliegen & Oxenham, 1999, Experiment 2). Nevertheless, Dowling (1973, Experiment 2) has shown with an immediate recognition task using unfamiliar melodies that a mean difference of 6 ST leads to recognition performance equivalent to that obtained for a one-octave difference if the melodies are composed of small pitch intervals. This result led him to formulate the hypothesis that melodic fission could depend on the degree of overlap of the frequency range covered by the melody and the distractor sequence. The results obtained by direct and indirect methods converge therefore to show the effect of frequency difference between pure tones on auditory stream formation, but they do not agree on the degree of difference needed to obtain segregation: 2–3 ST for an ABAB sequence to segregate and 6–12 ST to extract and recognize a melody interleaved with distractor tones or another melody. This divergence can probably be explained by the fact that in the case of interleaved melody recognition, it is a mean frequency difference, and not only a constant frequency difference, between two alternating tones, and also by the existence of context effects in streaming (Bregman, 1978b; Bregman & Rudnicky, 1975).

These different methods have also proven useful in demonstrating the role of other factors in the perceptual organization of sequences. A timbre difference between successive sounds—which is due to differences in spectral, temporal, or spectrotemporal properties of the sounds—induces perceptual fission in alternating sequences (Bregman et al., 1990; Cusack & Roberts, 2000; McAdams & Bregman, 1979; Singh, 1987; Singh & Bregman, 1997) and improves the recognition of an interleaved target pattern (Gregory, 1994; Hartmann & Johnson, 1991; Iverson, 1995; Wessel, 1979). An intensity difference also leads to perceptual fission of alternating sequences (van Noorden, 1977) and of interleaved melodies (Dowling, 1968; Hartmann & Johnson, 1991). Furthermore, differences in spatial position between sounds lead to the creation of distinct perceptual entities, as shown by the reinitialization of the cumulative effect of fission (Rogers & Bregman, 1998), and allow listeners to extract an intermixed melodic pattern when the sounds composing the pattern are not presented to the same ear as the distractor sounds (Deutsch, 1975, 1979; Dowling, 1968; Hartmann & Johnson, 1991).

The aim of the present study was (a) to develop a method that measures indirectly and objectively the primitive processes involved in auditory stream formation and (b) to establish a link between the results obtained with this indirect method, probing the precision of the representation of properties of a target stream (recognition performance) and those obtained with direct methods examining the perceptual fission phenomenon (fission boundary). We developed an interleaved melody postrecognition paradigm to this end. This experimental situation derives from the paradigm of immediate recognition of unfamiliar interleaved melodies proposed by Dowling (1973, Experiment 2). In his original paradigm, the listeners had to compare two successive melodies: a reference melody presented alone followed by a target melody interleaved with distractor notes. In our postrecognition paradigm, the reference (or probe) melody is presented after the composite sequence in order to minimize previous knowledge available to the listener concerning the target melody (Bey & McAdams, 2002). The aim is to maximize the involvement of primitive organization processes in the task.

After a preliminary experiment testing the paradigm, and in particular the algorithm used to construct the distractor sequences, three main experiments were performed. Experiment 1 examined the effects of a difference in mean (fundamental) frequency separating the target melody and the distractor sequence on postrecognition performance when the sounds were pure tones or six-harmonic complexes. The relation between postrecognition performance for interleaved melodies and the fission boundary was established, and an interleaved melody recognition boundary was defined. Experiment 2 studied the effect of timbre dissimilarity between target and distractor sequences on postrecognition performance. Finally, to ensure that the performance obtained reflected the immediate representation that the listener had of the composite sequence and not an a posteriori reconstruction (extraction of melodic cues after mental repetition of the sequence during the silent interval), the duration of the silent interval separating composite and probe melody sequences was varied in Experiment 3.

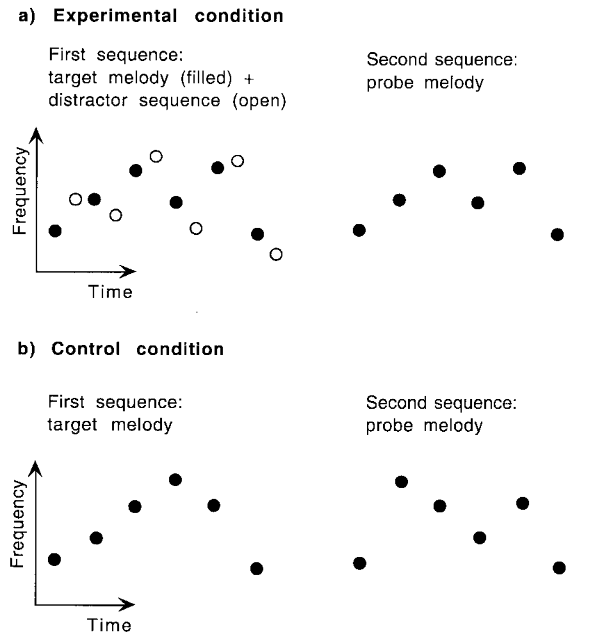

Figure 2. Visual illustration of the two types of conditions presented to listeners: (a) experimental condition in which the target melody is presented interleaved with a distractor sequence (target and probe melodies are identical in this example) and (b) control condition in which the melody is presented in isolation (target and probe melodies differ in the pitches of the second and fourth notes in this example).

General Method

An Interleaved Melody Postrecognition Task

Two sequences were presented in succession: (a) a mixed sequence composed of a six-note target melody interleaved with a six-note distractor sequence in the first interval and (b) an isolated six-note probe melody in the second interval. The probe melody was either identical to the target or differed in the frequencies of two of its notes. The listeners' task was to decide whether target and probe melodies were the same or different. The odd-numbered notes of the composite sequence formed the target melody, and the even-numbered notes the distractor sequence. The target and probe melodies were always presented in the same frequency register and with the same timbre; only the distractor sequence varied in mean frequency or in timbre, depending on the experiment. This experimental condition was followed by a control condition in which the target melody was presented without the distractor sequence. In this latter case, the task consists in a simple immediate melody recognition task (Dowling & Fujitani, 1971). These two conditions are illustrated in Figure 2. Postrecognition performance was compared with simple recognition performance in each experiment.

Stimuli

Thirty-six melodies and 180 distractor sequences, each with six notes, were created. The intervals were fixed, but the mean frequency of the sequences varied from trial to trial over a range from −3 to 2 ST.

Each of the 36 melodies had an original and a modified version (Appendix A). For the latter version, two notes—the second and fourth or the third and fifth—were changed within a range of ±4 ST. In all cases, the note changes altered the original melodic contour, that is to say, the direction of pitch change between successive notes. This feature is a salient cue for immediate, unfamiliar, diatonic, and nondiatonic melody recognition (Dowling, 1978; Dowling & Fujitani, 1971). These 72 melodies (36 original and 36 modified versions) were composed for the most part of ascending and descending pitch intervals, the size of which varied from 1 to 8 ST, but 7 of the modified melodies had repeated notes. All melodies were played within a one-octave range, their pitch ranges varying from 5 to 11 ST. The mean note was A5 (MIDI Note 81), with a fundamental frequency of 880 Hz for the experiments in which frequency difference varied (the preliminary experiment and Experiments 1 and 3) and Eb4 (MIDI Note 63 at 311 Hz) for Experiment 2 in which the segregation cue was timbre. Over the total set of 72 different melodies, 46 were diatonic and 26 nondiatonic. Diatonicity refers to the conformity of a melody to a diatonic scale, which corresponds to a specific pattern of intertone intervals in ST (e.g., the interval sequence 2 2 1 2 2 2 1 corresponds to a major scale). Note that the 46 diatonic melodies were not necessarily played in the same key, and that the strength of tonality (the sense of having a tonic reference pitch) varied across them as well. These two factors, key and tonality strength induced by a melody, were not studied systematically in this experiment. The original and modified versions were assigned in equal proportion to probe and target melodies.

Five distractor sequences were constructed for each of the 36 melody pairs (original and modified). They were all nondiatonic sequences with a pitch range varying from 7 to 15 ST. They were constructed with two constraints: (a) notes alternated from above to below the frequencies of the target melodies in order to create maximum crossover to camouflage the target (Hartmann & Johnson, 1991) and (b) the total range of the distractor sequence exceeded that of the targets at both upper and lower ends when the interleaved sequences were presented at the same mean frequency, with distractors being maximally distant from the two neighboring melody tones by 2 ST (Appendix B).

The sounds composing the sequences were 110 ms in duration. The interonset interval (IOI) was 165 ms for the composite sequences with 12 notes (6-note targets and 6-note distractors) and 330 ms for isolated melodies. The two sequences were separated by a 1,870-ms silence (except in Experiment 3).

Procedure

The experiments were conducted in three phases: familiarization, experimental, and control. Each test was preceded by eight familiarization trials of increasing difficulty during which feedback concerning the correct response was provided to participants. Participants responded by pressing one of two keys on a computer keyboard to indicate whether target and probe melodies were identical or different. The next trial followed automatically upon entering the response.

The melodies and corresponding distractor sequences were chosen randomly on each trial. In each frequency or timbre separation condition, an equal number of *same* and *different* trials were presented. In same trials, equal numbers of original and modified versions were assigned identically to target and probe melodies; in different trials, the order of assignment of original (o) and modified (m) versions to target and probe melodies was counterbalanced. Thus, all four trial types (oo, mm, om, mo) were equiprobable.

Analyses of variance (ANOVAs) were performed on arcsine transformed data (arcsine of the square root of proportion correct) to minimize potential problems of inhomogeneity of variance. The Geisser–Greenhouse correction (Greenhouse & Geisser, 1959) was applied to compensate for inhomogeneity of covariances due to repeated measures. F statistics are cited with uncorrected degrees of freedom. If epsilon (ε) is less than one, its value is cited, and the probability is determined with the corrected degrees of freedom.

Apparatus

The sounds were synthesized on a Yamaha TX802 FM tone generator. The synthesizer was controlled by a Macintosh SE/30 via a MIDI interface. The listeners were seated in a sound-treated room, and stimuli were presented over Sennheiser HD 520 II headphones connected directly to the output port of the synthesizer.

Preliminary Experiment: Algorithm Testing

We first tested the distracting power of each of the sequences interleaved with its corresponding target melody, to ensure that listeners can perform the task only when the composite sequence can be organized into two streams. We thus presented the distractor sequence in the same pitch range as the target melody (0-ST mean difference in frequency) to test if the listeners were able to do the task for the case in which the melody and distractor tones were perceptually grouped into one stream. Another condition in which the mean frequency separation between the melody and distractors was 24 ST (i.e., 2 octaves) was also presented to control for the ability of listeners to perform the interleaved melody recognition task when the melody and the distractors were clearly segregated (Dowling, 1973; Hartmann & Johnson, 1991). The experiment also allowed a verification of the equivalence of discriminability of the full set of target melodies. Indeed, the 36 melody pairs constructed have different characteristics in terms of melodic contour, position of changed notes, range of pitch change, and especially whether they conform to a diatonic musical scale, some being diatonic and others nondiatonic. These characteristics have been shown to be important in the perception and memory of melodies (Dowling & Harwood, 1986, chap. 5).

Method

Participants. Seventeen participants took part in the experiment, but 3 were removed from the analysis. Two listeners did not succeed in the interleaved melody recognition task; they performed virtually at chance level in all separation conditions ($M = .54$). One listener had absolute pitch perception and reported using a strategy to perform the task that consisted of memorizing the names of notes of the composite sequence, extracting every other note starting with the first one and then comparing this list with that of the probe melody. Our interest being to establish the relation between stream formation and interleaved melody perception, we decided to exclude listeners with absolute pitch because their performance did not depend only on their ability to perceptually separate the target melody and the distractor sequence. Therefore, the results of 14 listeners, 9 women and 5 men, were included in the analysis. Their ages ranged from 17 to 36 years ($M = 26.5$). Eight of them had played a musical instrument for at least 3–4 years.

Stimuli. All sounds were pure tones in this experiment. Five different distractor sequences (Versions a, b, c, d, and e) were constructed according to the same principle (Appendix B) for each of the 36 melody pairs. Two mean frequency separations between target melody and distractor sequence were presented: 0 ST and 24 ST. For the 24-ST separation, the mean frequency of the distractor sequences was lower than that of the target melody, which was always presented in the same register.

Procedure. Five experimental conditions testing the five sets of distractor sequences were presented to listeners in a counterbalanced order. This was followed by the control condition consisting of a simple melody discrimination task. These six tests were run on six different days. Each session lasted about 1 hr with a pause in the middle. Each experimental condition comprised 216 trials. The four possible trial structures (oo, mm, om, mo) were presented for each of the 36 melody pairs with the 0-ST separation (144 trials). For the 24-ST separation, only two trial structures were presented (one same and one different) for each given melody pair and a given listener in order to reduce the duration of the experiment (72 trials). However, the four trial types were presented with equal probability over the set of participants and were chosen randomly for each participant. The control condition comprised 144 trials. For each condition, an equal number of same and different trials was presented.

Results

Mean proportion of correct responses was computed over 14 listeners for the three conditions (0 ST, 24 ST, and control). When the target melody and distractor sequence were presented in the same frequency region, performance was near chance (correct proportions from .48 to .56; $M = .52$, $SD = \pm.02$). A detailed analysis of mean percentage of correct responses for the 180 combinations of distractors and targets revealed that all combinations at the 0-ST separation gave rise to virtually chance performance. When the mean distractor sequence frequency was 24 ST below that of the target melody, performance ($M = .94$, $SD = \pm.06$) approached that obtained in the control condition ($M = .97$, $SD = \pm.03$) but remained slightly lower. A two-way ANOVA with repeated measures on condition (0 ST, 24 ST, and control) and musical training (musicians and nonmusicians) as a between-subjects variable, with arcsine transformed proportions correct as a dependent variable, revealed the significant effect of conditions, $F(2, 24) = 216.80$, $\varepsilon = .88$, $p = .0001$. Fisher's least significant difference (LSD) post hoc comparisons confirmed that the 24-ST condition was much higher than the 0-ST condition ($p = .0001$) but significantly lower than the control condition ($p = .04$). This latter result suggests that even if the listener succeeds in perceptually segregating the target melody, the presence of the distractor sequence, even in a distant frequency region, slightly affects melody recognition. No effect of musical training was found, $F(1, 12) = 1.97$, $p = .19$.

Melodic discrimination errors in the control and 24-ST conditions were rare (3% on average with *SD* of 1%) and distributed in similar fashion across the set of melodies.

Discussion

The performance levels obtained in this interleaved melody postrecognition task thus seem to depend on the way listeners perceptually organize the composite sequence, with the exception of absolute pitch possessors (see the *Participants* section). Performance was at chance when the composite sequence would normally be heard as a single stream and near that obtained in a simple melody recognition task when the composite sequence would normally be heard as two streams. The results demonstrate the efficacy of the distractor construction algorithm in "camouflaging" the target melodies. Moreover, we found that the different characteristics of the melodies used, in particular the fact that some were diatonic and others nondiatonic, did not induce differences in their discriminability. This may be explained by the fact that the melody discrimination was relatively easy, as witnessed by the high performance in the control task. Several cues were in fact made available to the listeners to compare the melodies: two out of six notes were changed, the contour was modified, and, at times, even the diatonicity varied.

Experiment 1: Interleaved Melody Postrecognition on the Basis of Frequency Difference

The aim of Experiment 1 was to study the effects of a difference in mean (fundamental) frequency separating the target melody and the distractor sequence on postrecognition performance. Performances at different degrees of frequency separation (0, 6, 12, and 24 ST) are thus compared with the fission boundary, as measured with direct methods (Miller & Heise, 1950; van Noorden, 1975). The notion of partial melodic fission and the interleaved melody recognition boundary are proposed to establish a link between results obtained with this indirect method with interleaved melodies and those obtained previously with direct methods on alternating pure-tone sequences.

The sequences were played with pure tones and complex tones composed of six equal-amplitude harmonics in order to examine the effect of the spectral composition of the sounds on interleaved melody postrecognition as well. We thus tested whether a greater spectral overlap between the complex sounds of the target melody and distractor sequence would affect melodic fission. Some authors consider that the perceptual fission of successive sounds is due to the stimulation of distinct auditory filters (Anstis & Saida, 1985; Beauvois & Meddis, 1991; Hartmann & Johnson, 1991; van Noorden, 1975). According to this peripheral conception of stream formation, the stimulation of the same filters should thus reduce perceptual fission. The degree of overlap of the activity of the filters being greater when the sequences are played by complex sounds rather than by pure tones, the difference in mean (fundamental) frequency necessary to extract the target melody should consequently be larger for the complex sounds.

Method

Participants. Twenty-one listeners participated in the experiment. The performances of five listeners were excluded from the data analysis. One person reported having hearing problems after having performed the experiment. Four others did not succeed in the task when the distractor sequence was present. Their mean correct recognition rate averaged across all frequency separation conditions was .50 (.44–.55). Nevertheless, they were able to perform the melody discrimination task in the control condition without distractor tones: the mean correct recognition rate was .83 (.72–.91). The data were thus analyzed for 16 listeners, 10 women and 6 men, whose ages varied from 23 to 52 years ($M = 29$). Ten of these listeners had been playing a musical instrument for at least 3 years.

Stimuli. Four degrees of mean frequency separation were tested (0, 6, 12, and 24 ST). The distractor sequence was always transposed toward lower frequencies. The sounds composing the sequences were either pure tones or harmonic complexes composed of the first six harmonics with equal amplitudes. The loudness of the two sequence types was equalized globally by both of the authors. The maximum physical level achieved by each sound was measured with a Bruel & Kjaer 2209 sound-level meter (A-weighting, fast response) and was found to be about 76 dB for the pure tones and 67 dB for the complex tones (51.4 dB per component).

Procedure. The listeners performed two experimental conditions: one each with pure and complex tones in counterbalanced order and then the control condition with pure tones for half the listeners and complex tones for the other half. Each of these three tests comprised 32 trials (random selection of 32 melodies from among the 36). Each melody and an associated distractor sequence (chosen randomly from among the 5) were thus presented no more than once in each test. The session lasted about 30 min.

Results

A two-way ANOVA with repeated measures on tone type (pure and complex) and frequency separation (0, 6, 12, and 24 ST) revealed that performance for pure-tone stimuli was equivalent to that for complex-tone stimuli, $F(1, 15) = 0.59$, $p = .45$, and that this variable did not interact with the degree of frequency separation, $F(3, 45) = 0.51$, $\varepsilon = .83$, $p = .64$.

Because of the absence of an effect of tone type, data were combined for pure and complex tones for each degree of frequency separation. As can be seen in Figure 3, postrecognition performance increased with frequency separation. Performance was at chance when the target melody was presented in the same frequency region as the distractor sequence, and continued to improve for a separation greater than an octave (12 ST) reaching a level at 24 ST (.91) that approached that in the control condition (.98), although it remained slightly below this latter condition. For a correct proportion rate of .75, we can consider that the melody is recognized 50% of the time (chance performance = no recognition of the melody; perfect score = recognition of the melody in 100% of the cases; .75 is the middle of the scale). Therefore, the melodies were recognized in more than 50% of the cases when the distractor sounds were at an average frequency separation of more than 10 ST. Note that this value was computed by a linear interpolation (Figure 3). The separation corresponding to this performance level is called the interleaved melody recognition boundary (IMRB).

A one-way ANOVA with repeated measures on the four frequency separation conditions and the control condition revealed a highly significant effect of condition, $F(4, 60) = 53.14$, $\varepsilon = .81$, $p = .0001$. Fisher's LSD post hoc comparisons revealed significant differences between 0 ST and 6 ST ($p = .01$), 6 ST and 12 ST

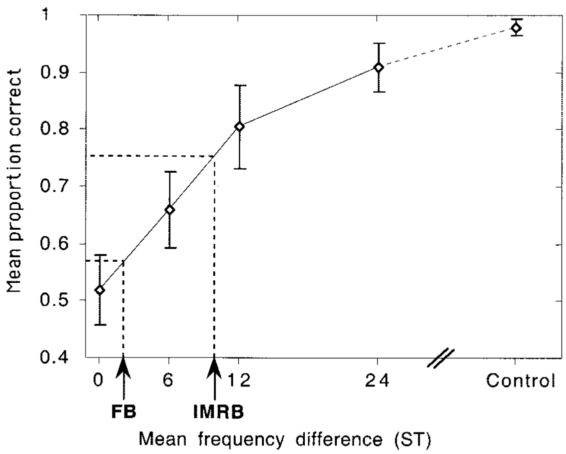

Figure 3. Experiment 1: mean proportion of correct responses for different frequency separation conditions (0, 6, 12, and 24 ST) and the control condition. The fission boundary (FB) for cyclical sequences (Miller & Heise, 1950; van Noorden, 1975) and the interleaved melody recognition boundary (IMRB) are indicated. Vertical bars represent 95% confidence intervals. ST = semitone.

($p = .001$), 12 ST and 24 ST ($p = .01$), as well as between 24 ST and the control condition ($p = .01$).

The performance of musicians was globally higher than that of nonmusicians. A two-way ANOVA, with repeated measures on frequency separation and with a between-subjects variable of musical training, showed a significant effect of musical training, $F(1, 14) = 7.10$, $p = .02$. The superiority of musicians in this task did not interact with the degree of frequency separation, $F(3, 42) = 0.88$, $\varepsilon = .88$, $p = .45$. Fisher's LSD post hoc comparisons revealed that the musicians' performance was significantly higher than that of nonmusicians only for 12 ST ($p = .01$) and 24 ST ($p = .04$). Performance on the control condition was comparable for the two groups. If we suppose that musicians can perform this simple melodic discrimination task better than nonmusicians, the equivalence of their observed performance could be explained by a ceiling effect because both groups obtained high scores.

The effect of musical training suggests that performance in the postrecognition task could be affected by previously acquired knowledge. We then tested to what extent performance could vary with learning. Because the listeners ran in the test two times, once for each tone type in a counterbalanced order, we examined whether performance improved in the second run averaged over tone types. The proportion of correct responses had a tendency to increase from the first run (.71) to the second run (.74). A three-way ANOVA on arcsine transformed proportion correct with repeated measures on frequency separation and run number and a between-subjects variable of musical training revealed that the effect of learning just barely failed to reach significance, $F(1, 14) = 4.54$, $p = .051$. Dowling (1973, Experiment 2) had also observed that immediate recognition of unfamiliar interleaved melodies had a tendency to improve over two experimental sessions.

Discussion

The improvement of postrecognition performance with increasing mean frequency separation between target and distractor sequences shows that this task allows for the measurement of primitive organization processes that are sensitive to signal characteristics. However, we found that a greater spectral overlap between the target and distractor sounds does not seem to affect melody recognition, suggesting that activation overlap on the basilar membrane is not determinant for stream formation, as has been convincingly demonstrated by Vliegen and Oxenham (1999) and in contradistinction to predictions by peripheral channeling models (e.g., Beauvois & Meddis, 1991).

Thus, increasing mean frequency induces progressive perceptual fission of the sounds composing the target melody, allowing the listener to access its melodic properties and compare it to the probe melody. We found that the melodies were recognized in 50% of the cases when the distractor sounds were at an average frequency separation of 10 ST. An IMRB of 10 ST was also found by Dowling (1973, Experiment 1) for other melodies and distractors. This melodic fission presents certain particularities that distinguish it from the fission of alternating pure-tone sequences. Indeed, a melody is, by definition, composed of more than two different pitches and, thus, of more than one pitch interval (in our study, the number of intervals can be as high as five). Consequently, the same mean frequency separation can induce the fission of two sounds composing an alternating sequence, while provoking only a partial fission of a melody interleaved with a distractor sequence, because of the presence of frequency intervals between target and distractor that are smaller than the mean frequency difference. Previous work has shown that a frequency difference of 2–3 ST suffices to lead to the formation of two single-frequency streams for alternating sequences (Miller & Heise, 1950; van Noorden, 1975). In the present study, if a linear interpolation was performed, listeners' performance improved at this frequency difference compared with the 0-ST condition and no longer responded at chance (Figure 3). This result suggests that more cues were available to listeners for 2–3 ST than for 0 ST, suggesting that melody was probably partially segregated beyond this frequency separation. The fission boundary would thus seem to correspond to a partial melodic fission.

Performance continued to improve beyond the IMRB and even beyond a mean separation of 12 ST, which was greater than the maximal range of our melodies. According to Dowling (1973), interleaved melody recognition may depend on the degree of overlap between the range of the target melody and the competing melody; that is, the separation between the lowest note of the target and the highest note of the distractor. It turned out that the range of our distractor sequences could extend to 15 ST in some cases (see the *Stimuli* section in General Method). This fact may explain why performance continued to improve at the 12-ST separation.

In spite of this improvement, performance at the 24-ST separation remained below that of the control condition, as in the preliminary experiment. This result suggests that the presence of the distractor sequence interfered with the target recognition task. Dowling (1973, Experiment 2) also observed that recognition of unfamiliar interleaved melodies was less than that obtained in the absence of distractors (Dowling & Fujitani, 1971). This degradation in recognition performance in the presence of a simultaneously presented sequence may have been due to an effect of attentional distraction. The listeners manifested difficulties in focusing their attention on the target stream, their attention alternating between the target and the distractor such that the detection of a change in the target may have escaped them. However, the degradation may also be explained by an interference in memory caused by the presence of the distractor sequence during the encoding of the target melody (Deutsch, 1970). This attentional and/or mnemonic interference suggests that this interleaved melody postrecognition task, designed to measure primitive processes involved in auditory stream formation, also probably brings into play other processes. Indeed, to accomplish the task, listeners must not only organize the composite sequence into two streams, but also focus their attention on the target melody—that is, select one of the two streams that are present, access the properties of this stream (notes, melodic contour, rhythm, etc.), and store this information in working memory.

Experiment 2: Postrecognition of Interleaved Melodies on the Basis of Timbral Dissimilarity

Several studies have demonstrated the role of acoustic cues related to timbre in the formation of auditory streams. Differences in spectral, temporal, and spectrotemporal parameters underlying this multidimensional auditory attribute promote perceptual fission in cyclical sequences (Bregman et al., 1990; McAdams & Bregman, 1979; Singh, 1987; Singh & Bregman, 1997) and facilitate melodic fission (Gregory, 1994; Hartmann & Johnson, 1991; Iverson, 1995, Experiment 2; Wessel, 1979). Iverson (1995, Experiment 1) has shown in particular that dissimilarity judgments between two timbres are good indicators of the "metric" used by the auditory system to segregate sound streams apparently coming from different musical instruments.

In the present experiment, the effect of timbral dissimilarity between target melody and distractor sequence on postrecognition performance was tested. The hypothesis was that if dissimilarity judgments predict fission, recognition performance should increase with timbral dissimilarity. The results of the preceding experiment suggested that interleaved melody recognition may also depend on the listeners' ability to focus their attention on the target melody. If some timbres are more intrinsically salient than others, they would be maintained in focus when they carried the target melody, but would be more attentionally distracting when in the distractor sequence. We thus varied not only the degree of dissimilarity between timbres, but also counterbalanced their assignment to targets and distractors.

Method

Participants. Thirty listeners participated in the experiment. The results of 2 of them were removed. One did not succeed in the experimental task: His correct recognition rate averaged across all timbre separation conditions was .38. The other failed the control condition (.44). Note that this percentage of rejected participants (7%) was lower than that found in the previous two experiments (12% and 19%, respectively). The remaining group of 28 listeners comprised 9 women and 19 men, with ages ranging from 19 to 51 years ($M = 30$), of which 17 had been playing a musical instrument for at least 3–4 years.

Stimuli. Synthetic sounds imitating musical instrument timbres—the vibraphone (VBS), guitar (GTR), trombone (TBN), and bassoon (BSN)—were used. The indicated labels are used to refer to the sounds as a reminder that they are synthetic imitations. These sounds were chosen from among a larger set for which three-dimensional timbre spaces were found by McAdams, Winsberg, Donnadieu, De Soete, and Krimphoff (1995).

The target melody and distractor sequence were presented at the same mean fundamental frequency (E♭4 or 311 Hz). This mean fundamental frequency was chosen for this experiment because it was the one used in the previous timbral dissimilarity studies for which the distances between the timbres of the instruments to be used are known. The target melody was played by the same instrument throughout an experimental condition; only the instrument assigned to the distractor sequence varied. The four instruments chosen had timbres varying on three perceptual dimensions with differing acoustical properties: (a) a spectral dimension related to the spectral centroid (mean of the frequency components weighted by their amplitudes), often associated with "brightness," "brilliance," or "nasality"; (b) a temporal dimension related to the attack quality and characterized by the logarithm of the rise time in the energy envelope; and (c) a dimension related either to variations in the spectral envelope over time, called "spectral flux" by McAdams et al. (1995), or to the degree of irregularity in the spectral envelope (Krimphoff, McAdams, & Winsberg, 1994), depending on the study.

Four degrees of timbral separation were used. The distractor sequence had the same timbre as the target (S0) or was distant by one, two, or three "steps" in the space (S1, S2, and S3), if the timbres are considered on an ordinal scale of distance with respect to the target timbre. To test the symmetry of the timbral distance effect on melodic fission, four experimental conditions were designed in which the target melodies were played by each of the four instruments. As such, the degrees of separation for S1, S2, and S3 corresponded to different instrumental timbres, depending on the target instrument. For example, if the target was played by VBS, the distractor was played by VBS for S0, GTR for S1, TBN for S2, and BSN for S3. However, if the target was played by BSN, for the distractor the instruments in order would be BSN, TBN, GTR, and VBS. There were six different pairs of instruments. The distances between the timbres (as estimated at the single fundamental frequency E♭4 in the original study) were computed from the timbre space determined by McAdams et al. (1995; see Table 1). Because the relation between judged dissimilarity and model distance is achieved with a linear fit, these distances correspond to an interval scale.

These distances were obtained from judgments made within a larger set of 18 sounds, whereas the present study considered only 4 of them. However, Donnadieu (1997) has shown that relative judgments of dissimilarity seem to change little when placed in a reduced timbre set. Furthermore, the dissimilarity judgments were performed on pairs of sounds, and it is still in terms of pairs (melody and distractor) that the current experiment was conducted.

The loudness and subjective duration of the sounds were equalized at the mean fundamental frequency (E♭4) by varying the MIDI velocity parameter available in the synthesizer. This parameter controls the intensity, but also the spectrum of the sound, as a function of the speed with which a key is pressed on a keyboard. The subjective duration was adjusted to compensate for subjective variations due to the presence of different forms of temporal envelopes on the instrument notes. The maximum level attained

Table 1
Distances Computed From the Timbre Space

Timbre pair	Relative distance
VBS–GTR	2.9
TBN–BSN	4.1
GTR–TBN	6.7
VBS–TBN	7.1
GTR–BSN	7.3
VBS–BSN	8.2

Note. Spaces determined by McAdams et al. (1995). VBS = vibraphone; GTR = guitar; TBN = trombone; BSN = bassoon.

by each sound was 59, 57, 64, and 65 dBA for VBS, GTR, TBN, and BSN, respectively.

Procedure. Participants ran four experimental conditions: one for each target timbre and then the control condition with one of the timbres (chosen with equal probability across the pool of participants). Each condition was preceded by a familiarization phase as in the preceding experiments. Effects of rank and order were counterbalanced across listeners. Each timbre condition was presented in each position an equal number of times, and each timbre was followed an equal number of times by each other timbre in order to counteract a possible influence of focusing on a given timbre on the succeeding condition. Each of the five conditions comprised 32 trials (randomly chosen without replacement from the 36 melodies). All other aspects were identical to Experiment 1. The session lasted about 1 hr, and listeners were paid for their participation.

Results

A two-way ANOVA with repeated measures on target melody timbre and timbral separation was performed with arcsine transformed proportions correct as the dependent variable. Performance improved with the degree of timbre difference between target melody and distractor sequence, $F(3, 81) = 47.16$, $\varepsilon = .85$, $p = .0001$. It did not depend on the timbre of the target melody, $F(3, 81) = 1.21$, $\varepsilon = .83$, $p = .31$, for the main effect, and $F(9, 243) = 1.79$, $\varepsilon = .76$, $p = .09$, for its interaction with timbral separation.

There were six different pairs of instruments defining six relative distances in the timbre space (Table 1) and the same-timbre condition (distance of zero). Because of the absence of a global effect of target timbre, data were thus combined for each distance independently of the specific timbre playing the target melody in order to establish a link between postrecognition performance and timbral dissimilarity. Postrecognition performance increased with timbral distance between target melody and distractor sequence (Figure 4). For a small difference in the timbre space (one unit equals one numerical category on the Dissimilarity Rating scale from 1 to 9 used in the McAdams et al., 1995, study), postrecognition performance appeared to be above chance. The distance

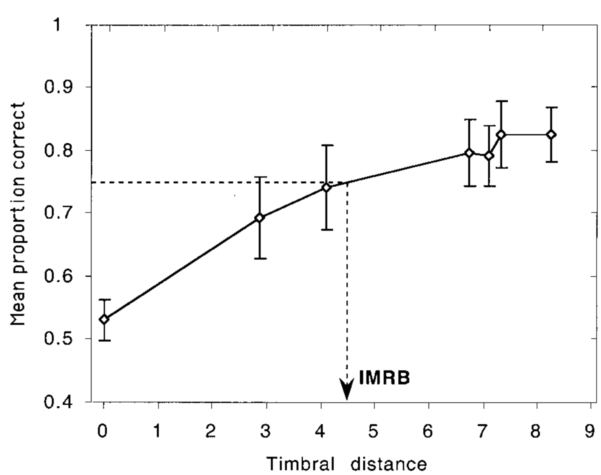

Figure 4. Experiment 2: mean proportion of correct responses as a function of timbral distance from the timbre space of McAdams et al. (1995) between instrument sounds playing the target melody and distractor sequence. Vertical bars represent 95% confidence intervals. IMRB = interleaved melody recognition boundary.

necessary to recognize the target melody 50% of the time (IMRB) was about 4.5 units on the rating scale by a linear interpolation.

A two-way ANOVA with repeated measures on timbral distance (seven timbre distances including zero; i.e., same-timbre condition) and a between-subjects variable of musical training was performed on arcsine transformed proportions correct. The effect of timbral dissimilarity was globally significant, $F(6, 156) = 15.97$, $\varepsilon = .77$, $p = .0001$. When the target and distractor were played by the same instrument, performance was near chance ($M = .53$, $SD = \pm.09$). When the timbre difference was maximal, the performance attained .83 ($SD = \pm.12$), which remains well below that for the control condition ($M = .95$, $SD = \pm.05$). As in the preceding experiments, the presence of a distractor sequence, even of very different acoustic properties, altered melody recognition compared with the case in which the melody was presented in isolation. No effect of musical training was found, $F(1, 26) = 1.02$, $p = .32$.

Discussion

Results of this experiment confirm that the degree of timbral dissimilarity, which takes into account the multidimensionality of this attribute, is a good predictor of melodic fission, extending the findings of Iverson (1995) on alternating-tone sequences. This suggests that dissimilarity judgments and the analysis performed by the auditory system to organize perceptually a sound sequence into auditory streams are based on a similar sensory representation. Moreover, we found that postrecognition performance did not depend on the target melody timbre, suggesting that performance in this task is a good reflection of the involvement of primitive processes.

The low failure rate and the absence of an effect of musical training, in contrast to the results from Experiment 1, suggests that this same task was easier to perform to some extent when the melodies were played by different types of sound sources than when they were played by simple synthetic sounds in different frequency registers. However, performance remained lower than that obtained for a mean frequency difference of 24 ST (see previous experiments). This could be due to the timbral dissimilarity used, which may not be equivalent to a difference of 24 ST for perceptual segregation. It could also be explained by a greater interference in memory when the distractor sequence was presented in the same pitch range as the melody in spite of the fact that it was not played with the same timbre (Semal & Demany, 1991, 1993).

Experiment 3: Effect of Retention Interval on Interleaved Melody Postrecognition

The preceding experiments have shown that postrecognition performance increases with mean frequency difference and timbral dissimilarity between target melody and distractor sequence. The results suggest that this task is in fact sensitive to primitive organization processes. However, it might be asked whether the measured performance reflects the immediate representation of the melody that emerges from the perceptual organization of the composite sequence or from an a posteriori reconstruction. Indeed, a listener engaged in the melody recognition task might try to play back mentally the first sequence during the time interval that separates the composite sequence and the probe melody in order to extract a posteriori the notes or other cues allowing the accomplishment of the task.

Experiment 3 examines the possible effect of mental repetition of the first sequence on postrecognition performance. The interleaved melody postrecognition task that was based on a frequency difference between target and distractor was presented to two groups of listeners. In the first group, the time interval between the composite sequence and the probe melody was reduced from 1,870 to 990 ms. This silent interval no longer allowed the listener to repeat mentally the composite sequence or the target melody at the original tempo. In a second group, the silent duration was further reduced to 330 ms so that mental repetition of the entire sequence was no longer feasible. Performance obtained from these two groups are compared with those collected from a subset of the participants in Experiment 2 in which the retention interval was 1,870 ms.

Method

Participants. Twenty-four participants, mostly university psychology students, took part in this experiment. Eight failed the test (i.e., a failure rate of 33%). Their performance averaged across all frequency separation conditions was .51 (.41–.63). Nevertheless, they were able to perform the melody discrimination task in the control condition without distractor tones: The mean correct recognition rate was .89 (.75–.97). The observed failure rate did not depend on the retention interval chosen (330 or 990 ms).

The data for the remaining 16 listeners were analyzed. The two groups were relatively homogeneous with respect to age, sex, and musical background. The mean ages of participants were 25 years (range: 22–27 years) with 7 women and 1 man for the 330-ms interval, and 25 years (range: 22–37 years) with 6 women and 2 men for the 990-ms interval. The number of musicians and nonmusicians was equal for each group. Four had played a musical instrument for at least 4 years, and the other 4 had never played an instrument. The results obtained for these two groups were compared with those of a third group drawn from the set of listeners having participated in Experiment 1 (ages varying from 23 to 52 years [$M = 31$] with 5 women and 3 men, 4 musicians and 4 nonmusicians) with a 1,870-ms interval. The participants selected for this analysis were chosen from among those that had passed the first condition in Experiment 1 with pure-tone stimuli (only those data are included in this analysis) and the control condition with either pure or complex tones.

Stimuli. The stimulus sequences were identical to those in Experiment 1. The effect of the three retention intervals (330, 990, and 1,870 ms) was tested in experimental and control conditions. The sequences were played with pure tones. For half of the listeners of the third group (from Experiment 1), the melodies in the control condition were played with complex tones composed of six equal-amplitude harmonics.

Procedure. The listeners performed the experimental condition and then the control condition, each composed of 32 trials (selected randomly from among the 36 melodies). Each condition was preceded by a familiarization phase as before. The experimental session lasted about 30 min.

Results

A two-way ANOVA with repeated measures on experimental condition (the four frequency separations and the control condition) and a between-subjects variable of retention interval was performed with arcsine transformed proportions correct. The reduction of the time interval separating composite sequence and probe melody had no differential effect on performance, $F(2, 21) = 0.50$, $p = .62$. Performance improved with increasing frequency difference as before, $F(4, 84) = 47.66$, $\varepsilon = .84$, $p = .0001$, but increased in the same fashion for all retention intervals, $F(8, 84) = 0.98$, $p = .45$.

Discussion

This experiment revealed that the decrease of the retention interval, avoiding the possible mental playback of the first sequence, did not affect the postrecognition performance. This result suggests that this paradigm measures the immediate representation of the melody that emerges from the perceptual organization of the composite sequence.

General Discussion

In this study, we have proposed an interleaved melody postrecognition task to measure indirectly auditory stream formation processes. This task is derived from the paradigm of interleaved melody recognition developed by Dowling (1973, Experiment 2). Contrary to the original paradigm, in the postrecognition task the melody to be recognized is presented after the sequence that must be perceptually organized in order to minimize the previously acquired knowledge that can be brought to bear in the perceptual analysis of the composite sequence.

Postrecognition performance increased with the difference in mean fundamental frequency (Experiments 1 and 3) and in timbral dissimilarity (Experiment 2) between the target melody and the distractor sequence. When the target and the distractor were presented in the same frequency register and played by the same instrument (having thus similar acoustical properties), performance was virtually at chance. When differences in frequency or spectrotemporal properties related to timbre increased, recognition performance improved and generally attained values near those for the control condition in which no distractor sequence was present. The melodies were recognized 50% of the time when the distractor sequence was distant on average by 10 ST, which we labeled IMRB. This value is higher than the fission boundary of cyclical sequences in which two pure tones alternate between different frequencies. However, at this latter threshold, which has been evaluated at 2–3 ST (Miller & Heise, 1950; van Noorden, 1975), performance was higher than in the 0-ST condition, and listeners did not perform at chance anymore. This suggests that the target melody was just starting to emerge from the composite sequence and was thus partially segregated. Timbre differences also led to melodic fission, and an IMRB of 4.5 units in the McAdams et al. (1995) timbre space was found. The effect of timbral dissimilarity on interleaved melody recognition performance confirms the fact that dissimilarity judgments constitute a good predictor of the degree of fission of a sequence (Iverson, 1995, Experiment 1). Finally, the third experiment revealed that the decrease of the retention interval, avoiding the possibility of mentally repeating the first sequence, did not affect postrecognition performance. This suggests that the interleaved melody postrecognition paradigm allows a measure of the immediate representation of the melody that emerges from the perceptual organization of the composite sequence.

This task was designed to minimize the intervention of previously acquired knowledge in the perceptual organization of a composite sequence. The melodies presented were not familiar to the listeners. The tests comprised few trials, thus limiting learning factors. The melodies were presented once or twice at most (with the exception of the preliminary experiment). Finally, the probe melody to be recognized was presented after the mixture. However, some knowledge, elaborated on during the experiment or related to the listener's expertise, may have intervened in this task. Indeed, some regularities specific to this experimental situation could be learned over the course of the session, such as the fact that the register of the target melody and its timbre remained constant (only the properties of the distractor sequence changed) and that in the experiments in which the frequency difference varied, the target melody was always presented in the higher register while the distractor moved toward the lower register. Dowling et al. (1987) have shown that pitch- and time-based expectancies can improve melody recognition. This kind of knowledge that the listener builds up over the course of the experiment may explain the tendency for performance to improve during the second run in Experiment 1 in spite of the small number of trials in each test (32 trials), a tendency also observed by Dowling (1973, Experiment 2) in an interleaved unfamiliar melody recognition task. The better results achieved by musicians (Experiment 1) also bears witness to the effect of learning, or at least learning skills, in this task. However, note that musicians improved their performance only for 12- and 24-ST conditions (i.e., when the melody was perceptually segregated from distractors).

Moreover, measuring auditory stream formation with an indirect method inevitably implies the involvement of other cognitive processes, such as attention and memory. In the postrecognition task, we found that performance in the maximal separation condition (in fundamental frequency or timbre) remained lower than that obtained in the control condition. This result was also observed by Dowling (1973; Dowling & Fujitani, 1971) and suggests that melody postrecognition does not depend only on the perceptual organization of the composite sequence into two streams. Indeed, to accomplish this task, listeners must not only organize the composite sequence into two distinct streams, but also focus their attention on the target stream, access the properties of the melody (contour, notes, intervals, or rhythm), and store this information in working memory. The degradation of recognition performance in the presence of the simultaneously presented distractor sequence can thus come from (a) an attentional distractability—the listeners manifesting difficulties in focusing their attention in a sustained fashion on the target stream such that the detection of a change in the target melody may escape them—and/or (b) an interference in memory caused by the presence of the distractor sequence (Deutsch, 1970), the simultaneous encoding of the distractor and the target being able in fact to degrade the representation of the target. This attentional and/or mnemonic interference indicates that in addition to the primitive organization processes, other cognitive processes intervene in the realization of this task.

Therefore, the involvement of other mechanisms invites caution in the inferences made on the number of streams perceived on the basis of performance in the recognition of a target melody. The preliminary experiment showed that with the exception of a listener who had absolute pitch, when the composite sequence was composed of interleaved melodies at the same mean frequency and designed for maximal contour crossing, listeners were not able to extract the target melody. Thus, the perception of a single stream implies an inability to perform this task. However, random performance does not imply, in turn, that the composite sequence was organized into a single stream. This is a possibility, but we cannot infer it with certainty because of the intervention of other processes in this task. Indeed, listeners could have perceptually segregated the melody and the distractor, but they had trouble with attentional focus or even problems with interference in working memory that altered their performance. This is probably what occurred in these experiments for the participants who failed to perform the task when the distractor sequence was presented even at large frequency differences. In the same manner, performance differences in the task induced by musical training (Experiment 1) do not necessarily reflect differences in organization, but probably the increased ability of musically trained listeners to focus on a melody in presence of other simultaneous tones. This observation means that there is not a perfect equivalence between the number of perceived streams and the fact that recognition performance is random or not. The recognition of a target pattern may depend on the perceptual organization of the composite sequence, as we have demonstrated for the interleaved melody postrecognition task. However, the perceptual analysis of this sequence does not guarantee the recognition of the target pattern, which also depends on other factors.

In conclusion, direct and indirect methods do not access the same representation. One is of a phenomenological nature measuring the perceptual experience the listener has of a change in analysis of the auditory scene, and the other examines the nature and precision of the representation of target stream properties. Therefore, with their different advantages and limits, these two complementary means of investigation reveal the different aspects of one general process: auditory scene analysis.

References

Alain, C., & Woods, C. A. (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.

Beauvois, M., & Meddis, R. (1991). A computer model of auditory stream segregation. *The Quarterly Journal of Experimental Psychology, 43A,* 517–541.

Bey, C., & McAdams, S. (2002). Schema-based processing in auditory scene analysis. *Perception & Psychophysics, 64,* 844–854.

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

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

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

Bregman, A. S. (1993). Auditory scene analysis: Hearing in complex environments. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 10–36). Oxford, England: Clarendon Press.

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

Bregman, A. S., Liao, C., & Levitan, R. (1990). Auditory grouping based on fundamental frequency and formant peak frequency. *Canadian Journal of Psychology, 44,* 400–413.

Bregman, A. S., & Rudnicky, A. (1975). Auditory segregation: Stream or streams? *Journal of Experimental Psychology: Human Perception and Performance, 1,* 263–267.

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

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

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

Deutsch, D. (1970). Tones and numbers: Specificity of interference in immediate memory. *Science, 68,* 1604–1605.

Deutsch, D. (1975). Two-channel listening to musical scales. *Journal of the Acoustical Society of America, 57,* 1156–1160.

Deutsch, D. (1979). Binaural integration of melodic patterns. *Perception & Psychophysics, 25,* 399–405.

Donnadieu, S. (1997). *Représentation mentale du timbre des sons complexes et effets de contexte* [Mental representation of the timbre of complex sounds and context effects]. Unpublished doctoral dissertation, Université Paris V, Paris.

Dowling, W. J. (1968). Rhythmic fission and perceptual organization. *Journal of the Acoustical Society of America, 44,* 369A.

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

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review, 85,* 341–354.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America, 49,* 524–531.

Dowling, W. J., & Harwood, D. L. (1986). *Music cognition.* Orlando, FL: Academic Press.

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

Greenhouse, S. W., & Geisser, S. (1959). On the methods in the analysis of profile data. *Psychometrika, 24,* 95–112.

Gregory, A. H. (1994). Timbre and auditory streaming. *Music Perception, 12,* 161–174.

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

Iverson, P. (1995). Auditory stream segregation by musical timbre: Effects of static and dynamic acoustic attributes. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 751–763.

Krimphoff, J., McAdams, S., & Winsberg, S. (1994). Caractérisation du timbre des sons complexes. II: Analyses acoustiques et quantification psychophysique [Characterization of the timbre of complex sounds. II: Acoustic analyses and psychophysical quantification]. *Journal de Physique, 4*(C5), 625–628.

MacDougall-Shackleton, S. A., Hulse, S. H., Gentner, T. Q., & White, W. (1998). Auditory scene analysis by European starlings (*Sturnus vulgaris*): Perceptual segregation of tone sequences. *Journal of the Acoustical Society of America, 103,* 3581–3587.

McAdams, S., & Bertoncini, J. (1997). Organization and discrimination of repeating sound sequences by newborn infants. *Journal of the Acoustical Society of America, 102,* 2945–2953.

McAdams, S., & Bregman, A. S. (1979). Hearing musical streams. *Computer Music Journal, 3,* 26–43.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research, 58,* 177–192.

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

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

Semal, C., & Demany, L. (1991). Dissociation of pitch from timbre in auditory short-term memory. *Journal of the Acoustical Society of America, 89,* 2404–2410.

Semal, C., & Demany, L. (1993). Further evidence for an autonomous processing of pitch in auditory short-term memory. *Journal of the Acoustical Society of America, 94,* 1315–1322.

Singh, P. G. (1987). Perceptual organization of complex-tone sequences: A tradeoff between pitch and timbre? *Journal of the Acoustical Society of America, 82,* 886–899.

Singh, P. G., & Bregman, A. S. (1997). The influence of different timbre attributes on the perceptual segregation of complex tone sequences. *Journal of the Acoustical Society of America, 102,* 1943–1952.

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

Sussman, E., Ritter, W., & Vaughan, J. H. G. (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology, 36,* 22–34.

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.

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

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

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

Wessel, D. L. (1979). Timbre space as a musical control structure. *Computer Music Journal, 3,* 45–52.

(*Appendixes follow*)

Appendix A

The Melodies

The 72 melodies constructed for the experiments, 36 original and their corresponding 36 modified versions, with two changed notes are shown in Table A1. The six notes of each melody are expressed in semitones with respect to the equal-tempered note the closest to the mean frequency of the original melodies. The value zero was assigned to A5 (MIDI Note 81) in the preliminary experiment and in Experiments 1 and 3 and to E♭4 (MIDI Note 63) in Experiment 2.

Table A1
The Full Set of 36 Original and Modified Melodies

Melody no.	Original melodies						Modified melodies					
1	2	4	2	1	−1	−3	2	0	2	−2	−1	−3
2	−3	−1	1	2	4	2	−3	−1	0	2	1	2
3	−2	0	−2	0	2	3	−2	0	−1	0	5	3
4	−2	0	2	3	2	3	−2	0	5	3	0	3
5	3	5	3	1	0	−4	3	1	3	−3	0	−4
6	−2	0	2	3	2	−2	−2	0	1	3	−2	−2
7	−3	−1	1	−1	1	4	−3	3	1	−2	1	4
8	3	5	3	1	−2	−4	3	2	3	−2	−2	−4
9	−2	0	2	3	0	−2	−2	0	4	3	−4	−2
10	−3	−1	−3	−1	3	4	−3	0	−3	3	3	4
11	−3	−1	1	−1	3	4	−3	−1	−3	−1	5	4
12	−1	1	3	1	−2	−1	−1	2	3	−3	−2	−1
13	3	5	3	0	−2	−4	3	5	1	0	−6	−4
14	−4	−2	0	3	5	3	−4	−2	3	3	2	3
15	−3	−1	−3	1	3	4	−3	−1	0	1	7	4
16	1	3	1	−2	0	1	1	0	1	−4	0	1
17	−1	1	3	−1	−2	−1	−1	5	3	1	−2	−1
18	−3	−1	1	4	3	4	−3	−1	3	4	6	4
19	1	3	5	1	0	−4	1	3	2	1	−3	−4
20	3	5	1	0	−2	−4	3	5	3	0	−6	−4
21	−2	0	3	2	0	−2	−2	2	3	−2	0	−2
22	−4	−2	2	3	5	3	−4	−2	6	3	7	3
23	−3	−1	3	4	3	4	−3	3	3	2	3	4
24	−4	−2	1	3	5	1	−4	−2	−1	3	2	1
25	−1	1	−3	−1	3	4	−1	1	0	−1	1	4
26	2	4	1	0	−3	−2	2	4	−1	0	0	−2
27	−3	−1	2	4	1	2	−3	−1	6	4	3	2
28	−2	2	3	2	0	−2	−2	2	1	2	−4	−2
29	−4	0	2	3	5	3	−4	−2	2	7	5	3
30	−3	1	−1	1	3	4	−3	−2	−1	2	3	4
31	−1	3	1	−1	−2	−1	−1	3	−3	−1	0	−1
32	−3	1	3	4	3	4	−3	1	6	4	5	4
33	−4	0	1	3	5	1	−4	3	1	2	5	1
34	−3	1	2	4	1	2	−3	4	2	0	1	2
35	−4	0	1	5	3	1	−4	0	5	5	2	1
36	−2	2	3	0	2	3	−2	2	4	0	5	3

Appendix B

The Distractor Sequences

To ensure that melody recognition performance depended on the perceptual organization of the first sequence composed of the target melody interleaved with a distractor sequence, it was necessary to construct the distractors so that listeners could not easily develop another strategy to perform the task, such as extracting cues related to the target melody contour from the global contour of the composite sequence. An algorithm was developed to generate five unique distractor sequences for each pair of original and modified melodies.

The algorithm determines the highest and lowest notes between notes i and $i + 1$ of the melody being considered. It randomly chooses the note j for the corresponding distractor to be 1–2 ST above the highest or 1–2 ST below the lowest melody note. Because each melody has two versions, the algorithm considers three possible cases: (a) the notes i and $i + 1$ are identical for both versions, (b) note i is different between the versions, or (c) note $i + 1$ is different. In the latter two cases, it makes its distractor choice on the basis of the highest and lowest of the three notes. If at choice n, note j of the distractor is in high position (above those of the melody pair), it will alternate to low position (below those of the melody pair) for choice $n + 1$, creating alternating ascending and descending intervals.

Therefore, when the target melody and distractor have the same mean fundamental frequency: (a) the crossing of the target melody pair by the distractor sequence was maximized, which is an important factor in the ability of a distractor sequence to camouflage the target melody (Hartmann & Johnson, 1991); (b) the target range was nested within the distractor range to avoid a contour-based judgment strategy because pilot experiments showed that when the target and distractor sequences were presented in the same register, and the probe melody differed by two notes, listeners at times achieved recognition performance above chance; and (c) the successive melody and distractor tones were maximally distant by 2 ST corresponding to the fission boundary (van Noorden, 1975) to ensure that the composite sequence will be perceived as one stream.

Because of the large number of different distractors, five (Versions a, b, c, d, and e) for each of the 36 melody pairs (a total of 180 sequences), a single example is given in Table B1 (melody pair and five distractors), and the melody pair and one distractor are illustrated in Figure B1.

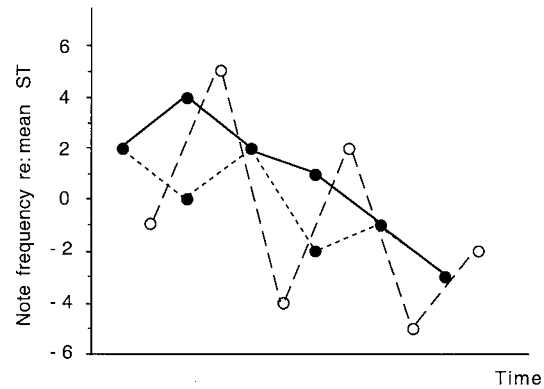

Figure B1. Graphic representation of the notes of Melody 1 (filled circles) in its original (solid lines) and modified (dashed lines) versions. The note frequency is expressed in semitones (ST) relative to the mean frequency of the melody. The interleaved Distractor Sequence a for this melody pair is shown with open circles. Note that the distractor notes alternate between high and low frequencies and always encompass the notes of the two versions of the melody.

Table B1
Example of Distractor Melodies Constructed for Target Melody 1 and Its Modified Version

Version	Interval pattern					
a	−1	5	−4	2	−5	−2
b	5	−1	3	−3	1	−4
c	−2	5	−3	3	−4	−2
d	6	−2	4	−3	1	−4
e	6	−2	3	−4	1	−4

Note. Distractor a is shown with the melody pair in Figure A1.

Received August 3, 2000
Revision received February 15, 2002
Accepted July 15, 2002 ■