The influence of different timbre attributes on the perceptual segregation of complex-tone sequences[a),b)]

Punita G. Singh[c)] and Albert S. Bregman
Department of Psychology, McGill University, Montréal, Québec H3A 1B1, Canada

(Received 29 September 1995; revised 15 May 1997; accepted 21 May 1997)

Spectral factors such as differences in harmonic content are powerful cues in the perceptual organization of tone sequences. Temporal features such as rise time, however, have been shown to be poor cues [W. M. Hartmann and D. Johnson, Mus. Perc. **9**, 155–184 (1991)]. The relative influence of these timbral features on perceptual segregation was investigated. Complex tones were sequenced in a repeating ABA- ''gallop'' format, under four conditions in which tones A and B had the same or different timbres as defined by differences in numbers of harmonics and temporal-envelope features. A sequence started with A and B tones at the same $F0$. The $F0$ difference between A and B then increased over the course of a trial, until a listener terminated the trial indicating perceptual segregation into sub-sequences comprising A and B tones, respectively. The $F0$ difference required to reach this crossover point of segregation provided a measure of the efficacy of stimulus features of A and B as cues for perceptual organization. Sequences combining differences in harmonic structure and temporal envelope required the smallest $F0$ change for segregation. Sequences of tones with the same harmonic structure and temporal envelope required larger changes in $F0$, while the other conditions fell in the middle of this range. The $F0$-tracking method used in this study facilitates measurement of the relative contribution of different stimulus features to stream segregation. It also holds potential as a tool using the point of segregation as a measure of the magnitude of timbre differences brought about by different physical features of sounds. [S0001-4966(97)03309-2]

PACS numbers: 43.10.Ln, 43.75.Cd, 43.66.Jh, 43.66.Lj, 43.66.Mk [WJS]

INTRODUCTION

The perceptual organization of sound sequences is dependent on several factors (Bregman, 1990; Deutsch, 1982; Handel, 1989; Jones, 1976; McAdams and Bregman, 1979). Some factors that influence how sounds will be perceptually organized in a sequential context are the range of $F0$ of the sounds, differences in spectral content and spatial location, and temporal proximity to other sounds. A sequence comprising pure tones that differ in frequency, for example, will at the appropriate tempo and frequency difference be perceived as splitting into sub-sequences within each of which the range of frequency differences is reduced.

A number of studies have investigated this perceptual segregation phenomenon using pure tones. Miller and Heise (1950) called the point of perceptual splitting the ''trill threshold'' while Dowling (1968) referred to the phenomenon as ''rhythmic fission.'' Bregman and Campbell (1971) called the perceptual splitting of a sequence into subsequences ''stream segregation.'' Van Noorden (1975) made a distinction between ''fission,'' the state when a sequence seems to be perceptually split into overlapping subsequences, and ''temporal coherence,'' the state in which the elements of a sequence remain perceptually integrated in a single sequence.

Different paradigms have been used to study stream segregation. For sequences of pure tones, frequency changes are generally correlated with changes in pitch. Experimental tasks have thus often been designed to utilize perception of pitch relations. For example, in some experiments, listeners were asked to identify melodies, the notes of which were interleaved so that the input sequence was a composite of the component melodies (Dowling, 1968; Hartmann and Johnson, 1991). Correct identification of the melodies would imply that listeners had perceptually segregated the input sequence into streams corresponding to the individual melodies.

Temporal-order perception is also affected when an input sequence is perceptually segregated (Hirsh, 1974). The streams typically appear to overlap in time, making it difficult to judge the actual order of elements in the sequence (Dannenbring and Bregman, 1976). This striking aspect of the streaming phenomenon has also been used in experimental tasks to determine the occurrence of perceptual segregation (Bregman and Campbell, 1971).

In addition to changes in perceived pitch and temporal-order relations, the rhythmic percept associated with a sequence can also be a powerful cue indicating perceptual segregation or coherence. A paradigm employed by van Noorden (1975) and in the present experiment, illustrates this phenomenon quite effectively. Three-element sequences

---

[a)] ''Selected research articles'' are ones chosen occasionally by the Editor-in-Chief, that are judged (a) to have a subject of wide acoustical interest, and (b) to be written for understanding by broad acoustical readership.

[b)] Some of this research was presented at a meeting of the Acoustical Society of America in Ottawa [Singh and Bregman, J. Acoust. Soc. Am. **93**, 2363 (A) (1993)].

[c)] Present address: 20-A Aurangzeb Road, New Delhi-110011, India.

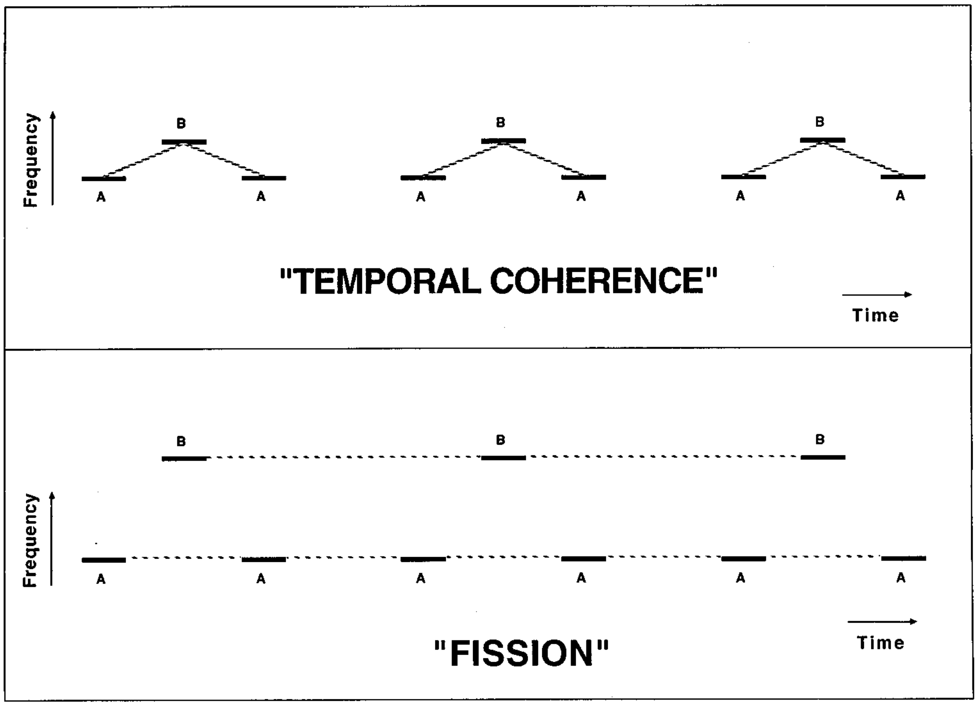


FIG. 1. The ''ABA-'' galloping sequence used by van Noorden (1975). The top panel represents retention of the galloping rhythm, in the ''temporal coherence'' state. The bottom panel illustrates the ''fission'' state, when the ABA- sequence perceptually segregates into overlapping sequences of A and B tones, with one sequence perceived to have double the tempo of the other.

composed of two tones A and B are created to form triplet patterns ABA as shown in Fig. 1. The ABA triplets are repeated with a gap equal to the duration of B inserted between repetitions. In the ''temporal coherence'' state, a sequence comprising such triplets appears to have a galloping rhythm. In the ''fission'' state, however, the sequence breaks up into perceptual streams comprising the A and B tones, respectively. Because of the temporal placement of the tones in the sequence, the sub-sequences will be isochronous, with the A tones perceived to repeat at a tempo twice that of the B tones. This dramatic change in rhythm is a useful cue indicating perceptual segregation.[1]

While the study of stream segregation for pure-tone sequences is relatively straightforward, for sequences of complex tones, the situation is more complex. Complex tones may differ from each other along several dimensions simultaneously. Thus tones with the same $F0$ may have very different spectra or onset-attack features. Would the type of segregation effects observed for sequences of pure tones hold for sequences comprising complex tones as well? If the sounds making up a sequence were produced by different instruments, would the formation of streams be dependent on the similarity of instrumental timbres or on the proximity of pitch as is typically the case for pure tones?

A number of studies have shown that stream segregation based on timbre differences is not only possible, but often more potent than segregation based on pitch differences (Iverson, 1993, 1995; Singh, 1987; Wessel, 1979). However, the manner in which stimulus features are manipulated to create timbre differences seems to be a crucial factor determining segregation. Spectral differences are particularly powerful initiators of stream segregation. A sequence comprising sounds differing in spectral loci of components will typically break up into streams, within which the range of spectral differences is reduced, even if the sounds share the same $F0$ (van Noorden, 1975; Singh, 1987). Timbre changes resulting from temporal differences, such as in attack and decay characteristics, however, have not proven to be very effective initiators of segregation (Hartmann and Johnson, 1991; Wessel, 1979).

Hartmann and Johnson (1991) investigated the influence of a variety of stimulus characteristics on stream segregation. Differences in amplitude-envelope shape were found to make no significant contribution to stream segregation. The relative dominance of spectral factors over temporal-envelope features was attributed to gross differences in peripheral channeling caused by spectral changes, and the absence or reduction of such differences given changes in envelope features alone.

''Peripheral channels'' in their report imply physiological channels that are involved in the initial stages of auditory processing based on frequency (tonotopicity) or on ear of presentation (laterality). According to their viewpoint, tones exciting different peripheral channels will be more likely to segregate from each other than those exciting the same channels. In the absence of differences in peripheral channeling, however, Hartmann and Johnson declare that little or no stream segregation will be observed, ''even in those cases where individual tones should clearly evoke images of different sources'' (p. 155).

At odds with the prediction of Hartmann and Johnson (1991), Iverson (1993, 1995) found an effect of envelope difference on stream segregation. Using edited samples of real instrument tones, Iverson found that, in addition to the usual spectral effects, sequences of tones with dissimilar temporal-amplitude envelopes received higher segregation ratings than sequences of tones with similar envelopes. Furthermore, tones with shorter attacks received higher ratings than tones with gradual attacks.

The conflict between the findings of Iverson and those of Hartmann and Johnson may be a result of the different stimuli used in the two studies. In most natural instrument sounds, the spectral and temporal dimensions may covary (Risset and Wessel, 1982). The separate contribution of these dimensions to stream segregation may therefore be difficult to ascertain. As admitted by Iverson (1993, p. 88), some ''unquantified acoustic attribute'' may have been correlated with the attack-time measures used in his study. Isolating the dynamic attributes that influenced streaming in his study clearly requires additional experiments using synthetic tones.

With synthetic sounds created in the laboratory, one can attempt to tease apart spectral and temporal dimensions and control them as independent variables in a streaming experiment. This was the intention of the present study. In particular, we wanted to determine the relative efficacy of differences in amplitude-envelope features, and harmonic content on stream segregation. A second goal was to devise a paradigm that would provide a common scale against which to measure or ''titrate'' the potency of different physical features as initiators of stream segregation. To obtain such a common measure, we used a variant of the van Noorden galloping ABA- sequence described earlier.

I. METHOD

A. Stimuli

Stimulus sequences were constructed following an ABA- format similar to that illustrated in Fig. 1. However, they were unlike those used by van Noorden (1975) in that the A and B sounds were complex, rather than pure tones, and could thus differ from each other along different timbral dimensions in addition to pitch (as defined by $F0$). Tones A and B were selected to have the same or different timbre as defined by similarity or difference in spectral and temporal features described below. Thus monotimbral AAA- sequences as well as bitimbral ABA- sequences were included in the stimulus inventory.

Both A and B were 100 ms in duration. A 10-ms silence was inserted between the 100-ms long tones, so that the physical onset-to-onset interval between tones A-B-A was 110 ms. A gap of 120 ms was inserted between repetitions of the ABA triplets. This type of temporal structuring leads to a characteristic galloping rhythm, that is lost when the sequence perceptually segregates into isochronous streams of A and B tones, respectively.

Sequences of sounds with different spectral and temporal-amplitude envelope features[2] were constructed following a two-factor design to generate four presentation conditions as summarized in Fig. 2.

The A and B tones could have either:

		Harmonics (n)	
		SAME	DIFFERENT
Envelope (e)	SAME	Condition 1 (Se Sn)	Condition 3 (Se Dn)
	DIFFERENT	Condition 2 (De Sn)	Condition 4 (De Dn)

FIG. 2. Sequences were constructed following the four conditions illustrated above. The A and B tones in the ABA- sequence could have either (1) the same envelope and number of harmonics (SeSn), (2) different envelope and same number of harmonics (DeSn), (3) same envelope but different number of harmonics (SeDn), or (4) different envelope and different number of harmonics (DeDn). Each condition had eight sub-conditions as described in the text and in Table AI.

(1) the same temporal envelope and same number of harmonics (SeSn)[3];
(2) different envelopes but same number of harmonics (DeSn);
(3) same envelope, but different number of harmonics (SeDn); or
(4) different envelopes and different number of harmonics (DeDn).

The spectral and temporal differences in design between individual sounds used in the stimulus sequences are illustrated in Fig. 3. The spectral factor had two levels, with tones constructed to have either the first two harmonics of the requisite $F0$, or the first four harmonics, added in phase at equal amplitudes. The two levels of the temporal factor corresponded to differences in the extent of rise and fall times. In one case, the tones had a 5-ms linear rise time with a 95-ms linear decay time. In the other case, the temporal-amplitude envelope was reversed so that the tones had a 95-ms rise time and a 5-ms decay time.

For each of the four sound designs illustrated in Fig. 3, a set of 25 tones ranging in $F0$ from 262 to 524 Hz were

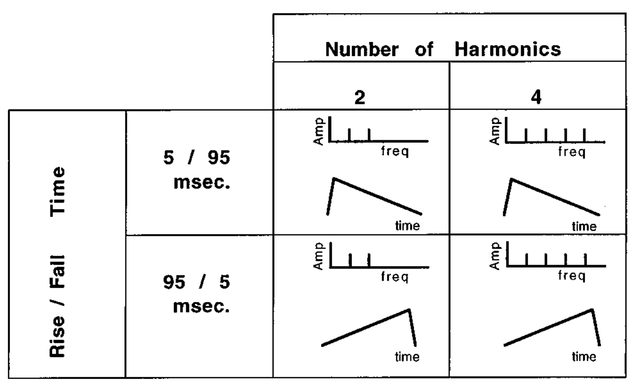


FIG. 3. Individual A and B sounds were constructed following a 2×2 design. The tones had either the first two or four harmonics of the required $F0$, and envelopes with either a sharp rise time and gradual fall time or a gradual rise time and sharp decay time as illustrated above.

synthesized. The 25 tones spanning this octave interval differed in $F0$ from each other in quarter-tone steps, where a quarter-tone step is equivalent to a change in $F0$ by half a semitone, i.e., by an amount equal to $2^{1/24}$ (or 3%) of the reference $F0$. All sounds were equalized in rms energy to compensate for the difference in the number of components. The entire set of tones was accessible via the control program described below for selection during the adaptive procedure used.

B. Apparatus

All sounds were synthesized digitally using the MITSYN software package for signal processing and analysis (Henke, 1990). Sound synthesis, stimulus presentation and data collection were controlled by a 486/50 microcomputer fitted with a Data Translation DT 2823 audio card set at a sampling frequency of 20 kHz and 16-bit resolution. The stimuli were filtered via TTE low-pass filters set at a cutoff frequency of 8 kHz with a 96 dB/oct roll off. Output presentation levels were controlled via a Tascam amplifier/mixer and verified with a GenRad 1565 sound level meter. The listener was seated in an IAC double-walled, sound-absorptive booth and received the stimuli binaurally via Sennheiser HD414 headphones at an overall sound-pressure level of about 70 dB.

C. Procedure

The stimuli were presented in an interactive procedure constructed using the MAPLE software package (Achim *et al.*, 1992). On each trial, a listener was presented with a repeating ABA sequence in which the $F0$ difference between A and B was initially 0 Hz. As the trial proceeded, the $F0$ difference between A and B increased in quarter-tone steps, following an ascending track (*re*: 262 Hz) or descending track (*re*: 524 Hz), until the sequence appeared to perceptually segregate. At this ''crossover point,'' the trial was terminated by the listener pressing a key on the computer terminal and the amount of $F0$ change in quarter-tone steps was recorded.

Since attention has been shown to have an influence on perceptual segregation boundaries (van Noorden, 1975) listeners were specifically instructed to try to hold on to the galloping pattern despite the changes in $F0$. They were to terminate the trial only when the galloping rhythm was lost and they perceived the sequence to have segregated into streams.

The procedural control program monitored the selection of timbral features and fundamental frequencies of the A and B tones during a trial. For every four repetitions of the ABA-pattern, the $F0$ of the middle tone B changed so that the $F0$ interval between A and B accordingly changed by a quarter-tone. The direction of change of $F0$ within a trial (up or down) and the order of particular sounds serving as A and B in the sequence were counterbalanced so that each of the four main conditions had eight subconditions. Features of individual sounds used in the 32 subconditions thus resulting, are summarized in Table AI in the Appendix. The 32 stimuli were presented randomly in a block, with six replications obtained per subject.

D. Subjects

Ten listeners between the ages of 21 and 36 years were used as subjects. They all had normal hearing and had participated in auditory experiments before. All subjects were given a block of practice trials encompassing all stimuli used to familiarize them with the task prior to actual data collection. Individual results are described in Sec. II B.

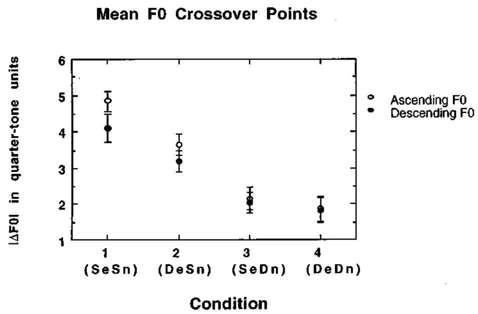
II. RESULTS

Mean $F0$ values obtained at segregation crossover points for all subjects across replications for the 32 stimuli were highest for condition 1, getting progressively lower for conditions 2, 3, and 4. A two-way analysis of variance (ANOVA), with condition and direction of $F0$ change considered as the two factors, revealed a highly significant effect for condition [$F(3,27)=59.29$, $p<0.000$],[4] a significant effect of $F0$ direction [$F(1,9)=6.79$, $p<0.027$], and a significant interaction between condition and $F0$ direction [$F(3,27)=7.21$, $p<0.001$]. One-way analyses designed to probe the interaction between condition and $F0$ direction revealed that the $F0$-direction factor was coming into play only in conditions 1 and 2 [$F(1,9)=11.62$, $p<0.007$ and $F(1,9)=6.65$, $p<0.029$, respectively]. For conditions 3 and 4, the $F0$-direction factor was not significant [$F(1,9)=0.6$, $p=0.467$ and $F(1,9)=0.77$, $p=0.754$, respectively].

To evaluate the effect of order of timbres of A and B in a sequence, the means corresponding to subconditions for conditions 2, 3, and 4 were analyzed via a three-way ANOVA with condition, $F0$ direction, and A-B timbre order considered as the three factors. Subconditions for condition 1 (SeSn) were not included in this analysis since the sequences for this condition were monotimbral, i.e., of an AAA- format, where timbre order was not an issue. As expected, the effect of condition was highly significant [$F(2,18)=33.40$, $p<0.000$]. The main effect of $F0$ direction, however, was not significant [$F(1,9)=3.12$, $p=0.109$] but the condition $\times F0$ direction interaction remained [$F(2,18)=4.31$, $p<0.029$]. No effect of order of timbres A and B was found [$F(1,9)=2.31$, $p=0.161$].

Condition 1 (SeSn) was also analyzed to determine if there was any effect of absolute features of the sounds such as steepness of the amplitude envelope or the number of harmonics on the crossover points for monotimbral AAA-sequences. This was done via a three-way ANOVA with two levels for each of the three factors under consideration, i.e.: direction of $F0$ change within a trial (up or down), steepness of rise time (5 or 95 ms) and number of harmonics (two or four).

The $F0$-direction factor proved to be highly significant [$F(1,9)=11.66$, $p<0.008$]. There was no effect of number of harmonics [$F(1,9)=1.61$, $p=0.235$] and of steepness of envelope [$F(1,9)=4.00$, $p=0.074$]. The interaction of en-



Condition Comparisons	Ascending F(1,9)	Trials p	Descending F(1,9)	Trials p
1 vs. 2	37.51	< 0.000	37.73	< 0.000
1 vs. 3	171.19	< 0.000	62.22	< 0.000
1 vs. 4	89.64	< 0.000	39.93	< 0.000
3 vs. 4	4.09	= 0.072	2.22	= 0.168

FIG. 4. Mean $\Delta F0$-crossover points (in quarter-tone steps) obtained for ten listeners for the four main conditions shown along the abscissa. Empty symbols correspond to crossover points for ascending $F0$ trials. Filled symbols represent descending $F0$ trials. Error bars correspond to the standard error of the data. Results of planned comparisons for different condition pairs are given at the bottom.

velope and $F0$ direction, however, was significant $[F(1,9) = 7.33,\ p < 0.023]$. The interaction was explored further via one-way analyses that revealed that $F0$ direction only made a difference for envelopes with the sharper rise times $[F(1,9) = 19.99,\ p < 0.002]$.

The main results of these analyses are summarized below:

(1) The effect of condition (1, 2, 3, or 4) was highly significant.
(2) Direction of $F0$ change made a significant difference to crossover points in conditions 1 and 2 but not 3 and 4.
(3) The order of timbres for A or B tones did not contribute to any significant differences in crossover points.
(4) Monotimbral AAA- sequences of sounds with envelopes with steep rise times (5 ms) had significantly higher $\Delta F0$ crossover points than sequences of sounds with more gradual rise times (95 ms) for ascending $F0$ trials.
(5) The absolute number of harmonics (two or four) in AAA- sequences did not make a difference to crossover points.

$\Delta F0$-crossover points averaged across listeners for the four main conditions of the experiment are shown in Fig. 4 for both ascending and descending $F0$ trials. Crossover points for individual listeners are listed in Tables I and II and discussed later. The ordinate in Fig. 4 gives the mean $F0$ difference between tones required for stream segregation in quarter-tone steps (where a quarter-tone is about 3% of the

TABLE I. Mean crossover points for the four main conditions for ascending $F0$ trials for individual listeners. Numbers in parentheses below the means are the respective standard errors. The last four columns of the table summarize the results of planned comparisons for conditions 1 vs 2, 1 vs 3, 1 vs 4, and 3 vs 4 for each listener. Overall means across subjects are also given at the bottom of the table for comparison.

Subject	Condition 1 (SeSn)	Condition 2 (DeSn)	Condition 3 (SeDn)	Condition 4 (DeDn)	1 vs 2 p	1 vs 3 p	1 vs 4 p	3 vs 4 p
S1	4.08	3.08	1.62	1.33	<0.057	<0.000	<0.001	=0.057
	(0.23)	(0.27)	(0.08)	(0.14)				
S2	4.67	3.58	2.62	2.62	<0.001	<0.003	<0.002	=1.000
	(0.31)	(0.26)	(0.15)	(0.21)				
S3	5.04	3.46	2.17	1.37	<0.002	<0.000	<0.000	<0.004
	(0.14)	(0.22)	(0.14)	(0.08)				
S4	6.33	4.17	3.54	2.46	<0.001	<0.002	<0.000	=0.057
	(0.31)	(0.22)	(0.42)	(0.18)				
S5	3.21	2.12	1.00	1.00	<0.005	<0.000	<0.001	=1.000
	(0.27)	(0.21)	(0.20)	(0.13)				
S6	4.42	3.21	0.92	0.75	<0.004	<0.001	<0.000	=0.469
	(0.34)	(0.26)	(0.25)	(0.09)				
S7	5.58	3.58	2.25	1.62	<0.000	<0.000	<0.000	<0.007
	(0.14)	(0.17)	(0.09)	(0.18)				
S8	4.96	4.50	1.29	0.79	=0.099	<0.001	<0.000	=0.166
	(0.30)	(0.20)	(0.18)	(0.21)				
S9	5.71	5.54	3.71	4.08	=0.855	<0.024	=0.174	=0.762
	(0.74)	(0.81)	(0.94)	(0.68)				
S10	4.42	3.25	2.46	2.62	<0.043	<0.002	<0.011	=0.589
	(0.26)	(0.44)	(0.27)	(0.29)				
Mean	4.84	3.65	2.16	1.87	<0.000	<0.000	<0.000	=0.072
	(0.28)	(0.29)	(0.31)	(0.34)				

TABLE II. Same as Table I, but for descending $F0$ trials.

Subject	Condition 1 (SeSn)	Condition 2 (DeSn)	Condition 3 (SeDn)	Condition 4 (DeDn)	1 vs 2 p	1 vs 3 p	1 vs 4 p	3 vs 4 p
S1	3.71 (0.18)	3.12 (0.14)	2.25 (0.13)	1.92 (0.17)	<0.013	<0.001	<0.002	=0.248
S2	3.21 (0.10)	2.87 (0.08)	2.37 (0.15)	1.96 (0.15)	=0.081	<0.005	<0.001	<0.030
S3	3.62 (0.29)	2.79 (0.23)	1.17 (0.10)	0.96 (0.15)	<0.030	<0.001	<0.001	=0.091
S4	7.00 (0.22)	4.92 (0.25)	3.54 (0.18)	2.67 (0.08)	<0.002	<0.000	<0.000	<0.008
S5	3.00 (0.17)	1.79 (0.22)	1.12 (0.19)	0.87 (0.17)	<0.000	<0.000	<0.000	=0.142
S6	3.29 (0.16)	2.67 (0.27)	1.25 (0.26)	0.79 (0.21)	<0.010	<0.001	<0.000	=0.193
S7	4.29 (0.08)	3.33 (0.10)	1.96 (0.15)	1.58 (0.12)	<0.003	<0.000	<0.000	=0.059
S8	4.50 (0.24)	3.58 (0.31)	1.29 (0.27)	0.87 (0.14)	<0.009	<0.000	<0.000	<0.041
S9	5.21 (0.41)	4.21 (0.44)	3.54 (0.51)	4.04 (1.09)	<0.042	<0.024	=0.285	=0.629
S10	3.29 (0.15)	2.58 (0.21)	2.00 (0.30)	2.67 (0.26)	<0.038	<0.016	=0.108	=0.161
Mean	4.11 (0.39)	3.19 (0.28)	2.05 (0.29)	1.83 (0.33)	<0.000	<0.000	<0.000	<0.168

reference $F0$). The abscissa shows the corresponding conditions as defined in Fig. 2.

A. General observations

Figure 4 shows a declining trend for crossover points across conditions 1–4 for both ascending and descending $F0$ trials. Crossover points for ascending trials were higher than those for descending trials for conditions 1 and 2, but not for conditions 3 and 4. The reason for this difference is not clear at present.

For both ascending and descending trials, the highest average $\Delta F0$ values were obtained for condition 1 (SeSn), in which A and B tones were designed to have the same timbre (i.e., the same temporal envelope and number of harmonics). Condition 1 (SeSn) serves as a reference condition with no changes in harmonic numbers or temporal envelopes across tones of the sequence. Planned comparisons between condition 1 and the other three conditions showed significant differences as summarized at the bottom of Fig. 4. Crossover points for condition 2 (DeSn), were significantly lower than those for condition 1 [ascending, $F(1,9)=37.51$, $p<0.000$; descending $F(1,9)=37.73$, $p<0.000$]. In this condition, A and B differed in envelope, but had the same number of harmonics. Crossover points were even lower for condition 3 (SeDn), in which A and B had the same envelopes, but differed in harmonic structure, and lowest for condition 4 (DeDn) in which A and B differed in both envelope and harmonic content.

The $\Delta F0$ values obtained for condition 3 were significantly lower than those for condition 2 [ascending $F(1,9)=36.86$, $p<0.000$; descending $F(1,9)=39.05$, $p<0.000$]. A difference in harmonic structure alone was thus more powerful in facilitating segregation of A and B than an envelope difference alone. Supplementing an envelope difference by a difference in harmonic numbers led to even further lowering of crossover points as is evident from a comparison of conditions 2 vs 4 [ascending, $F(1,9)=41.84$, $p<0.000$; descending $F(1,9)=22.89$, $p<0.001$]. Supplementing a harmonic difference with an envelope difference, however, did not lead to significant lowering of crossover points for condition 4 as contrasted with condition 3 [ascending, $F(1,9)=4.09$, $p=0.072$; descending $F(1,9)=2.22$, $p=0.168$].

Improved segregation for harmonic differences (condition 1 vs 3 and 1 vs 4) is not surprising, given the growing body of evidence implicating spectral differences as enhancers of stream segregation. The significant difference between condition 1 and 2, however, is contrary to the prediction of Hartmann and Johnson (1991) who found no effect of envelope differences on stream segregation. This may be constructed as supporting the case of Iverson (1993, 1995) for segregation based on onset differences, but it should be noted that the difference between conditions 3 and 4, which also differed from each other only in the envelope parameter, was not statistically significant.

B. Individual differences

Crossover points for individual listeners for the four main conditions of the experiment are presented in Tables I and II for ascending and descending $F0$ trials, respectively. These $\Delta F0$ values were obtained by averaging cross the six replications for each listener. Standard errors are given in parentheses below the means. The mean results for all ten listeners are also given for comparison in the last row of the tables. The last four columns of each table summarize the results of planned comparisons of different conditions in terms of the level of statistical significance.

As can be seen from the tables, *absolute* $\Delta F0$ values were quite different for different listeners. Listener S4 for example, could hold an ascending-trial sequence together under condition 1 for 7 quarter-tone units on the average ($\approx$21%). Listener S5 on the other hand, achieved segregation at 3 quarter-tone units ($\approx$9%). However, *relative* differences between conditions showed the same declining trend as the mean data. Conditions 2, 3, and 4 were significantly different from the standard condition 1 (SeSn) for most listeners. However, the difference between condition 3 and 4 which was not statistically significant for the listener-averaged crossover values was statistically significant for some listeners.

Subjects in our experiment were not preselected on the basis of musical experience. However, different degrees of familiarity with music may have contributed to some of the individual differences observed. Pitt (1994) has observed that nonmusicians are more sensitive to changes in timbre than to changes in pitch in sound categorization tasks. Musicians on the other hand, tend to follow pitch relations more closely. Cho *et al.* (1994) also note that familiarity with particular instrument timbres may affect the relative weighting assigned by listeners to physical features of sounds.

In the present experiment, such factors may have contributed to differences in absolute values of crossover points. In reference to Tables I and II, listener S1 sings in a choir, listener S4 is an accomplished pianist, listener S5 plays the saxophone, and listener S8 (the first author) is a percussionist. Despite musical exposure being a common factor, the absolute values of crossover points for these listeners are different. However, as noted above, relative values show the same trend across these and other listeners.

To systematically evaluate the effect of musical training on stream segregation, experiments would have to be done in which subjects were selected not only on the basis of general musical experience *per se*, but also based on different types of musical experience.

III. GENERAL DISCUSSION

The results of the present experiment corroborate the importance of spectral differences in facilitating stream segregation. Conditions 3 (SeDn) and 4 (DeDn), under which the sounds in a sequence differed from each other in terms of number of harmonics, led to significantly lower $\Delta F0$ values for segregation than the reference condition 1 (SeSn), in which there was no difference between sounds in terms of harmonic structure. However, our listeners also attained significantly lower crossover points for condition 2 (DeSn), under which changes were made only in the envelope feature of sounds. In this condition, the sounds comprising the sequence occupied the same peripheral channels at unison, differing only in the time course of evolution of amplitude. Their long-term power spectra were identical.

Though the mean crossover points for conditions 1 (SeSn) and 2 (DeSn) were significantly different, those for conditions 3 (SeDn) and 4 (DeDn) were not. The reason for the difference between these complementary conditions is not clear. It could be that for conditions 3 and 4 the $\Delta F0$ values were approaching ''floor'' limits. The harmonic structure difference alone was large enough to cause streaming without much change in pitch. The influence of the added envelope difference may thus not have been observable at this low end of the scale.

One could also speculate that in condition 2 (DeSn), the lack of a concurrent difference along the harmonic dimension allowed the envelope differences to be better detected. These perceived envelope differences were apparently adequate to enhance stream segregation so that segregation was achieved at lower $F0$ differences than the null standard. For condition 3 (SeDn), the spectral differences alone were a highly effective cue for segregation. The additional difference provided by a change in envelope in condition 4 (DeDn) apparently did not serve to enhance perceptual segregation of A and B any further.

A study by Grey (1978) investigating timbre discrimination in musical patterns suggests that sequences camouflage the temporal detail of individual sounds while amplifying spectral differences between sounds. Isolated contexts, on the other hand, appear to facilitate the comparison of temporal features of a pair of tones, such as differences in their rates of attack and decay (p. 471). The salient spectral difference between sounds used in condition 4 (DeDn) may have had such an obscuring effect on the concurrent envelope difference for our stimuli.

In the present study, onset-to-onset times between sounds were not adjusted for individual listeners to compensate for *perceptual* attack times of sounds with different envelopes. The perceptual correlate associated with envelope changes, while producing a timbre change, may also have provided a slight rhythmic cue (perceived as a difference in ''accent'' within the sequence). Any advantage obtained via this rhythmic cue, however, is observed only for condition 2 (DeSn) versus condition 1 (SeSn), not condition 4 (DeDn) versus 3 (SeDn), which should have similarly benefitted from this cue.

Iverson (1993, 1995) investigated the influence of such inadvertent rhythmic cues in a comparative study using sequences with elements spaced by physically equal onset-to-onset intervals as well as sequences in which the spacing of elements was adjusted in an attempt to provide equal perceptual onset-to-onset intervals. No difference between these two types of sequences was found in terms of stream-segregation measures. It thus seems unlikely that accent structure affected crossover points in our study. Furthermore, such an effect would have shown up as an A-B timbre order effect in the statistical analyses described earlier, but was not observed.

Iverson (1993, 1995) found a correlation between dissimilarity measures of sounds and measures of stream segregation. He suggested that auditory stream segregation is based on the same dynamic and spectral acoustic attributes that influence similarity judgments. The same factors that allow listeners to discriminate sounds, should help in segregating them in a sequential context.

It is reasonable to assume that some differences must exist between sounds in a sequence, for subgroups such as streams to emerge. But how different must these differences be, in order to be successful initiators of stream segregation? Added to the problem of obtaining adequate measures of magnitude of difference, is that of having a common way of measuring the effects of differences along various stimulus dimensions.

With the $F0$-tracking method used in the present experiment, we were attempting to provide such a measure that would allow comparisons of different stimulus features in terms of their contribution to the $\Delta F0$ segregation value. The method was successful for this purpose, in that it enabled a ranking of conditions in terms of crossover points in common, underlying $F0$-change units.

In the present experiment, stimulus features were varied following the design table in Fig. 3. Tones corresponding to the four recipes illustrated clearly differed from each other in perceived timbre. Although timbre discrimination *per se* was not the goal of our study, discrimination experiments conducted by other investigators indicate that the type and magnitude of stimulus features manipulated by us should evoke discriminable changes in timbre (Samson *et al.*, 1993). The relative contribution of different physical features in evoking a perceived change in timbre, however, may be different. For example, Cho *et al.* (1993, 1994) found spectral factors to be more crucial to the normalization of instrument timbre than temporal factors such as attack times. In addition to differences in relative weighting of physical substrates of timbre, the perceived magnitude of timbre difference between sounds may also differentially affect stream segregation.

If the degree of perceived dissimilarity of timbres is indeed a predictor of stream segregation (Iverson, 1993, 1995), then the difference in crossover point in $F0$ units would also provide a measure of the discriminability of timbres contrasted in the sequence. Sequences of sounds with very different timbres would be likely to segregate at lower $\Delta F0$ values, while those with more subtle variations in timbre would segregate at higher $\Delta F0$ values. The relative values of crossover points would be an indicator of the degree of perceived difference between timbres despite different physical substrates contributing to the timbre difference. The method used here could thus potentially be used to obtain quantitative measures of timbre difference.

One limitation of the $F0$-tracking method, however, relates to ranking of conditions such as 3 and 4 of the present experiment. For these conditions, the $\Delta F0$ crossover points appeared to be approaching limiting values at the lower end of the $\Delta F0$ scale. Indeed, it is possible to achieve stream segregation at unison ($\Delta F0 = 0$ Hz) for some types of timbre contrasts between sounds (Iverson, 1993, 1995; Singh, 1987). This suggests development of an analogous tracking procedure based on tempo manipulation rather than frequency manipulation for sequences contrasting highly distinct timbres. Rate of presentation of sounds in a sequence is also one of the key factors bringing about segregation (van Noorden, 1975). Multitimbral sequences of complex tones at different fixed $\Delta F0$ values could be used as stimuli, and an adaptive procedure used with rate of presentation (ΔT) changing over the course of a trial. Listeners would terminate trials at ΔT values where segregation appeared to occur. Lower ΔT values (i.e., faster tempi) would presumably be required for sequences more resistant to segregation, while sequences comprising highly contrastive sounds more amenable to segregation would break apart at larger ΔT values (i.e., slower tempi). A wider range of timbre contrasts could thus be studied using this complementary procedure.

IV. CONCLUSIONS

The present study attempted to measure the relative efficacy of different stimulus dimensions for initiation of perceptual segregation of sequences. Sounds composing the sequences had the same or a different number of harmonics and temporal-amplitude envelopes. Maximal segregation was obtained for sequences that combined differences along both these dimensions, as characterized by low $\Delta F0$ values at the point of segregation. Differences in harmonic numbers were the next best at causing segregation. Monotimbral conditions in which the sounds in a sequence shared the same number of harmonics and envelope characteristics were the least susceptible to segregation (i.e., they had the highest $F0$ crossover points).

A surprising finding is that crossover points for sequences of sounds with envelope differences alone, though higher than the harmonic-number conditions, still proved to be significantly lower than the standard condition. Given the results of Hartmann and Johnson (1991), Iverson (1993, 1995), and the present study, the role of temporal cues in stream segregation clearly needs to be studied further. A wider range of envelope differences and temporal-envelope modulation differences could be included in future investigations.

The method used in the present study enabled measurement of the influence of different timbre attributes on perceptual segregation using $\Delta F0$ crossover points as a common denominator. The stimuli used were kept deliberately simple in the present experiment, but in future research, we hope to apply this adaptive procedure using complex stimuli with a wider range of spectral and temporal differences. With some streamlining in terms of specification of frequency stepsizes used, temporal points of change of $F0$, tempo of the sequence, etc., the method used in the present study holds promise for obtaining quantitative measures of stream segregation, as well as providing a way to measure differences in complex perceptual attributes such as timbre in common units.

TABLE AI. Summary of the features of the sounds used in construction of sequences ABA- for conditions 1, 2, 3, 4, and their subconditions.

	Tone A		Tone B		
Condition	Number of harmonics	Rise–fall time in ms	Number of harmonics	Rise–fall time in ms	$F0$ change direction
SeSn					
1.1	2	5/95	2	5/95	ascending
1.2	4	5/95	4	5/95	ascending
1.3	2	95/5	2	95/5	ascending
1.4	4	95/5	4	95/5	ascending
1.5	2	5/95	2	5/95	descending
1.6	4	5/95	4	5/95	descending
1.7	2	95/5	2	95/5	descending
1.8	4	95/5	4	95/5	descending
DeSn					
2.1	2	5/95	2	95/5	ascending
2.2	2	95/5	2	5/95	ascending
2.3	4	5/95	4	95/5	ascending
2.4	4	95/5	4	5/95	ascending
2.5	2	5/95	2	95/5	descending
2.6	2	95/5	2	5/95	descending
2.7	4	5/95	4	95/5	descending
2.8	4	95/5	4	5/95	descending
SeDn					
3.1	2	5/95	4	5/95	ascending
3.2	4	5/95	2	5/95	ascending
3.3	2	95/5	4	95/5	ascending
3.4	4	95/5	2	95/5	ascending
3.5	2	5/95	4	5/95	descending
3.6	4	5/95	2	5/95	descending
3.7	2	95/5	4	95/5	descending
3.8	4	95/5	2	95/5	descending
DeDn					
4.1	2	5/95	4	95/5	ascending
4.2	4	95/5	2	5/95	ascending
4.3	2	95/5	4	5/95	ascending
4.4	4	5/95	2	95/5	ascending
4.5	2	5/95	4	95/5	descending
4.6	4	95/5	2	5/95	descending
4.7	2	95/5	4	5/95	descending
4.8	4	5/95	2	95/5	descending

ACKNOWLEDGMENTS

Financial support for this project was provided by the Natural Sciences and Engineering Research Council of Canada. The authors are grateful to Pierre Ahad and Todd Mondor for technical and statistical assistance and to the editor and reviewers for their patience and suggestions.

APPENDIX

A summary of features of sounds used in construction of sequences ABA- for conditions 1, 2, 3, 4, and their subconditions is given in Table AI.

[1]An audio example of the type of sequences used by van Noorden (1975) can be heard on the compact disc of Auditory Demonstrations distributed by the Acoustical Society of America.

[2]The terms ''spectral'' and ''temporal'' are used here in an operational sense to define the stimulus variables ''harmonic structure'' and ''amplitude envelope,'' respectively. In reality, it is not possible to separate the spectral from the temporal completely. A ''temporal'' feature such as sharp attack time has a spectral consequence (in evoking high-frequency distortion in an output audio device) just as a ''spectral'' feature such as number of harmonics has a temporal consequence (in that the waveform becomes more complex).

[3]Note that for condition 1 (SeSn), the sequences are in effect monotimbral, i.e., of an AAA- type while for the other conditions, the sequences were bitimbral, i.e., of an ABA- type. In this study A and B are used as timbre labels and do not have any bearing on the $F0$ relation between the tones. Thus AAA- sequences do not imply tones of the same $F0$, but rather of the same timbre, as defined by similarity of harmonic structure, envelope, or both of these features.

[4]The statistical program used was set to round off values to the third decimal place. The use of the expression ''$p<0.000$'' here and elsewhere, actually implies that the probability of obtaining the result purely by chance would be lower than 5 in 10 000 (i.e., $p<0.0005$).

Achim, A., Bregman, A. S., and Ahad, P. (**1992**). MAPLE software documentation, Speech and Hearing Laboratory, Department of Psychology, McGill University, Montréal, Québec, Canada.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA).

Bregman, A. S., and Campbell, J. (**1971**). ''Primary auditory stream segregation, and the perception of order in rapid sequences of tones,'' J. Exp. Psychol. **89,** 244–249.

Cho, J. L., Hall, M. D., and Pastore, R. E. (**1993**). ''Stimulus properties critical to normalization of instrument timbre,'' J. Acoust. Soc. Am. **93,** 2402.

Cho, J. L., Hall, M. D., and Pastore, R. E. (**1994**). ''Normalization of musical instrument timbre,'' Center for Cognitive and Psycholinguistic Studies, SUNY, Binghamton, New York (unpublished).

Dannenbring, G. L., and Bregman, A. S. (**1976**). ''Stream segregation and the illusion of overlap,'' J. Exp. Psychol.: Hum. Perc. Perf. **2**, 544–555.

Deutsch, D. (**1982**). ''Grouping mechanisms in music,'' in *The Psychology of Music* (Academic, New York), Chap. 4.

Dowling, W. J. (**1968**). ''Rhythmic fission and perceptual organization,'' J. Acoust. Soc. Am. **44,** 369.

Grey, J. M. (**1978**). ''Timbre discrimination in musical patterns,'' J. Acoust. Soc. Am. **64,** 467–472.

Handel, S. (**1989**). *Listening: An Introduction to the Perception of Auditory Events* (MIT, Cambridge, MA).

Hartmann, W. M., and Johnson, D. (**1991**). ''Stream segregation and peripheral channeling,'' Mus. Perc. **9**, 155–184.

Henke, W. H. (**1990**). ''MITSYN: A coherent family of command-level utilities for time signal processing,'' available from WLH, 133 Bright St., Belmont, MA 02178.

Hirsh, I. J. (**1974**). ''Temporal order and auditory perception,'' in *Sensations and Measurement*, edited by H. R. Moskowitz, B. Scharf, and J. C. Stevens (Riedel, Dordrecht, The Netherlands), pp. 251–258.

Iverson, P. (**1993**). ''Auditory stream segregation by musical timbre,'' Ph.D. dissertation, Cornell University, Ithaca, New York.

Iverson, P. (**1995**). ''Auditory stream segregation by musical timbre: Effects of static and dynamic acoustic attributes,'' J. Exp. Psychol.: Hum. Pec. Perf. **21**, 751–763.

Jones, M. R. (**1976**). ''Time, our lost dimension: Toward a new theory of perception, attention and memory,'' Psychol. Rev. **82,** 323–355.

McAdams, S., and Bregman, A. S. (**1979**). ''Hearing musical streams,'' Comput. Music J. **3,** 26–43.

Miller, G. A., and Heise, G. A. (**1950**). ''The trill threshold,'' J. Acoust. Soc. Am. **22,** 637–638.

van Noorden, L. P. A. S. (**1975**). ''Temporal coherence in the perception of tone sequences,'' Ph.D. dissertation, IPO, The Netherlands.

Pitt, M. A. (**1994**) ''Perception of pitch and timbre by musically trained and untrained listeners,'' J. Exp. Psychol.: Hum. Perc. Perf. **20**, 976–986.

Risset, J. C., and Wessel, D. L. (**1982**). ''Exploration of timbre by analysis and synthesis,'' in *The Psychology of Music*, edited by D. Deutsch (Academic, New York), Chap. 2.

Samson, S., Zatorre, R., and Ramsay, J. (**1993**). ''Multidimensional scaling of synthetic musical timbre: Perception of spectral and temporal characteristics,'' J. Acoust. Soc. Am. **93,** 2402.

Singh, P. G. (**1987**). ''Perceptual organization of complex-tone sequences: A tradeoff between pitch and timbre?,'' J. Acoust. Soc. Am. **82,** 886–899.

Singh, P. G., and Bregman, A. S. (**1993**). ''Spectro-temporal factors in the perceptual segregation of tonal sequences,'' J. Acoust. Soc. Am. **93,** 2363A.

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