Available online at www.sciencedirect.com
SCIENCE DIRECT®
Hearing Research 193 (2004) 95–104
HEARING RESEARCH
www.elsevier.com/locate/heares

Effects of differences in the pattern of amplitude envelopes across harmonics on auditory stream segregation

Rhodri Cusack [a,b,*], Brian Roberts [a]

[a] School of Psychology, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK
[b] MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, UK
Available online 27 April 2004

Abstract

When a mixture of sound from many sources arrives at the ear, the auditory system attempts to segregate it into different perceptual streams. For sequences of sounds, the effects of basic acoustic properties (e.g., frequency separation, rate) on streaming are well understood, but much less is known about the effects of more complex acoustic attributes. Dynamic variations in frequency spectrum are known to have an important effect on the timbre of sounds. We investigated whether dynamic variations also affect stream segregation. Periodic tones were used, comprising harmonics 1–6, and presented in long sequences (ABA–ABA–···). Tones A and B always differed in fundamental frequency, a factor known to influence streaming, and had either the same or different patterns of spectral variation. The total amount of variation (spectral flux) was held constant. Listeners judged moment-by-moment the grouping of these sequences, and the measure used was the proportion of time a sequence was heard as segregated. We found that sequences of complex tones with different patterns of spectral variation are more likely to segregate when this results in different patterns of change in the frequency centroid over time (Experiments 1 and 2), but not when it does not (Experiment 3).

Keywords: Auditory stream segregation; Timbre; Complex sounds

1. Introduction

The sound arriving at the ear is often a mixture arising from multiple sources, so that the brain is faced with the problem of sorting out which components have arisen from which source. Fortunately, there is a range of regularities across sound sources that can be used *to organize perceptually* the auditory scene into separate streams (Bregman, 1990). For example, the sounds produced by a periodic source typically change in fundamental frequency (F_0) only relatively slowly over time. Descriptions of the scene analysis problem usually break down into two parts: that of separating simultaneously occurring elements from independent sources and that of forming links across time between elements arising from the same source. The current study pertains to this second problem of sequential grouping.

A sound usually evokes several perceptual attributes. An important perceptual attribute is loudness, which is primarily influenced by intensity. Another is pitch, which for harmonic sounds is primarily determined by F_0. Other possible perceptual differences are lumped together into the catch-all term ''timbre''. It is well established that the grouping of sequentially presented pure tones is governed primarily by frequency separation and by rate (e.g., Bregman and Campbell, 1971; van Noorden, 1975). It has also been shown that differences between complex sounds in spectral content and periodicity can affect their sequential grouping (e.g., Wessel, 1979; Singh, 1987; Bregman et al., 1990; Iverson, 1995; Cusack and Roberts, 1999, 2000; Roberts et al., 2002). This might be due to differences in the timbre that these sounds evoke (for a discussion, see Moore and Gockel, 2002). Alternatively, different factors might underlie similarity of timbre and stream segregation, in which

* Corresponding author. Tel.: +44-1223-355294x661; fax: +44-1223-359062.
E-mail address: rhodri.cusack@mrc-cbu.cam.ac.uk (R. Cusack).

Abbreviations: ERB, equivalent rectangular bandwidth; F_0, fundamental frequency; HHPE, higher harmonics peak earlier; IHPE, inner harmonics peak earlier; LHPE, lower harmonics peak earlier; OHPE, outer harmonics peak earlier

0378-5955/$ - see front matter © 2004 Published by Elsevier B.V.
doi:10.1016/j.heares.2004.03.009

case sounds that are more dissimilar in timbre might not necessarily show greater stream segregation.

In the following two subsections, we first summarize studies that have attempted to characterize timbre and then we discuss which acoustic characteristics have been shown to affect stream segregation.

1.1. Acoustic characteristics underlying timbre

In several studies, listeners were asked to rate pairs of complex sounds for similarity in timbre. These ratings were analyzed using multidimensional scaling (Shepard, 1962a,b; Kruskal, 1964a,b) in an attempt to identify the underlying factors that listeners use to make their judgements. Grey (1977), Wessel (1979), and Iverson and Krumhansl (1993) used recorded musical instrument sounds equalized for pitch and loudness. Krumhansl et al. (reported in Krumhansl, 1989) used synthesized musical sounds. One dimension identified in all of these studies was the brightness of a sound, which is related to its spectral center-of-gravity. Another dimension found in all of these studies was one related to characteristics of the attack, but there was some disagreement on the exact property that was important. A third dimension, thought to be related to the degree of synchrony between the harmonics of a sound, was identified in some of the studies (Grey, 1977; Krumhansl et al., see Krumhansl, 1989). At one end of this dimension, the sounds comprised components that underwent synchronous changes and maintained the same relative amplitudes throughout (i.e., no spectral flux). At the other end, the sounds had components that developed at different times, thus leading to changes in their relative amplitudes (i.e., high spectral flux).

Grey and Gordon (1978) modified half of the sounds in Grey's (1977) original set, by taking pairs of tones and exchanging the mean amplitudes of the corresponding harmonics within each pair, but preserving the original temporal envelopes of the harmonics. The other tones were presented unchanged. Once again, listeners were asked to rate the similarity of pairs of the tones and their judgements were analyzed using multidimensional scaling. A three-dimensional solution was found, similar to that of Grey (1977). Those tones with exchanged spectral energy distributions were found to have swapped places on the dimension attributed to brightness, but had remained in similar positions on the other two dimensions. This confirmed that the dimension corresponding to brightness was related to the spectral energy distribution, while the other two dimensions depended on the temporal variation of the stimuli.

In summary, a key finding has been that not only the steady-state spectrum, but also the variation of sounds over time makes an important contribution to timbre. In particular, there is good evidence of roles for some characteristic of the attack and for the overall pattern of spectral variation over time. It might then be expected that differences in these attributes will also affect stream segregation. Iverson (1995), Singh and Bregman (1997), and Cusack (1998) have all shown that differences in attack abruptness between tones can enhance their stream segregation. In the current study, the role of spectral variations over time was examined.

1.2. Acoustic characteristics affecting stream segregation

Iverson (1995) attempted to identify what acoustic characteristics have the greatest effect on stream segregation. He played a sequence of a pair of musical instrument sounds of the same pitch and loudness alternating repeatedly. Listeners were asked to rate the extent of segregation and these ratings were then analyzed in two ways. In the first method, multidimensional scaling was used in a procedure similar to that used in the studies of timbre described above. Two factors were found, which were interpreted as corresponding to brightness and attack abruptness. Iverson did not find a dimension corresponding to the pattern of spectral variation. In a second part of the analysis, Iverson calculated several measures of various acoustic features for all of the sounds. These included the following dynamic acoustic properties: the pattern of change over time in the frequency centroid (spectral center-of-gravity), the amount of spectral flux (degree of spectral change across time), and the sharpness of attack. Multiple regression was then performed to find, for each pair of tones, the extent to which these measures affected their segregation. Overall, this second analysis does provide some evidence that spectral variations in the sounds affect their stream segregation. Iverson's key findings can be summarized as follows:

(1) The degree of correlation of the temporal pattern of the frequency centroids of sounds in a sequence was a good predictor of stream segregation. Sounds with a more similar pattern of change were more likely to integrate into a single stream.
(2) Similarity between sounds in the amount of spectral flux did *not* influence their sequential grouping. However, there was greater stream segregation when the total amount of flux was high, averaged over all the sounds in a sequence.
(3) Sounds with more abrupt onsets (sharper attacks) tended to segregate from each other to a greater extent than did sounds with gentler onsets.

These findings can be related to the studies on timbre discussed previously. Grey (1977), Grey and Gordon (1978), and Krumhansl et al. (see Krumhansl, 1989) all suggested that the timbre of a sound was affected by the degree of similarity between the amplitude envelopes of its different harmonics. This is closely related to the amount of spectral flux, which becomes greater as the temporal envelopes of the individual components of a

sound become more dissimilar. In terms of differences between sounds that might influence their sequential grouping, Iverson (1995) found that the difference in the amount of spectral flux was not a good predictor of streaming. Note, however, that an increase in spectral flux is only one consequence of introducing differences between the amplitude envelopes of the components of a sound. Other effects may include dynamic changes in frequency centroid, in spectral compactness/dispersion, and in spectral roll-off. Indeed, two sounds may differ in any of these properties without necessarily differing in the amount of spectral flux.

Some caution is needed when interpreting judgements of recorded complex sounds that have been analyzed using multidimensional scaling or multiple regression. It can be argued that, although a correlation has been found between an acoustic characteristic and a particular effect on timbre or streaming, this attribute might not be the cause of the effect. There might be another attribute that was not investigated, which also correlated with the effect, and was the true cause. Such a co-occurrence of acoustic attributes might indeed be expected owing to the excitation and resonance constraints on sound sources (see Handel, 1989; Risset and Wessel, 1999). The sounds produced by musical instruments provide a number of examples of these correlated acoustic properties. For example, an increase in playing effort increases the bandwidth of a sound primarily by increasing energy in the higher harmonics, leading to a higher frequency centroid. This effect is particularly strong for brass instruments. Also, it is true for most instruments that the frequency centroid of a sound increases with F_0. Another example is bowing style on stringed instruments, which produces correlated changes in the sharpness of onset and in the spectral content of the attack. Hence, for example, it would be possible to misattribute a true effect on streaming of differences in bandwidth or F_0 to correlated differences in frequency centroid.

To avoid possible confounds owing to correlated acoustic attributes, simple and precisely controlled complex tones were used in the current study. The aim was to investigate whether or not differences in the *pattern* of spectral variations can affect stream segregation when differences in the *amount* of spectral flux are eliminated. The measure of frequency centroid used in the current study was computed from the amplitude envelopes of the harmonics on a linear frequency scale. It should be noted, however, that equating properties of sounds according to physical measures does not necessarily equate their perceptual correlates.

2. Experiment 1

Iverson (1995) concluded that stream segregation is influenced by the similarity of the pattern of variation in frequency centroid across individual sounds in a sequence. In Experiment 1, this was tested by measuring the streaming of sequences comprising sounds that were either the same or different in the pattern of variation of their centroids. There were no differences in the total amount of spectral flux in the sounds.

2.1. Method

2.1.1. Design

Listeners were asked to judge the extent of segregation of a repeating sequence in a task similar to that first used by Anstis and Saida (1985), where the perceived state is indicated throughout the trial. Instead of a simple alternating sequence, a more complex one similar to that of van Noorden (1975) was employed, so that differences in stream segregation led to a difference in the rhythm perceived. This combination of task and sequence structure has previously been found to provide a sensitive and reliable measure of stream segregation (Cusack and Roberts, 1999; Carlyon et al., 2001; Roberts et al., 2002). The extent of streaming was measured as a function of F_0 separation and for various combinations of temporal envelopes for the harmonics. The sounds were given the same total amount of spectral flux, but different patterns of variation. It was predicted that sounds with a more similar pattern of variation in their frequency centroid would show less stream segregation.

2.1.2. Participants

Eight undergraduate students participated in this experiment, which lasted about 1 h, for either payment or course credits. All reported normal hearing.

2.1.3. Stimuli

A 30-s repeating sequence of a complex tone with a lower F_0 (A), a complex tone with a higher F_0 (B), and a silent interval (–) was played (ABA–ABA–$\cdots$). This sequence, like that used in the stream segregation studies of van Noorden (1975), is heard as a distinctive ''galloping'' rhythm when the low and high sounds are heard as a single stream, but not when the sequence is heard as two streams. This change in rhythm provides a salient cue that makes the task easier. The tones and the silent interval were each 100 ms in duration, so that each complete cycle took 400 ms. The first six harmonics were present in all the sounds. Each harmonic rose linearly in amplitude over time, reached a peak, and then decreased linearly in amplitude back down to zero. All of the harmonics began and ended at the same time.

There were two different patterns of temporal variation for the harmonics of the tones. In one pattern, the peak time for harmonic n was $10 + 16(n - 1)$ ms, giving 10, 26, 42, 58, 74, and 90 ms for the six harmonics, respectively. This pattern is called ''lower harmonics peak earlier'' (LHPE), and the amplitude envelopes for the

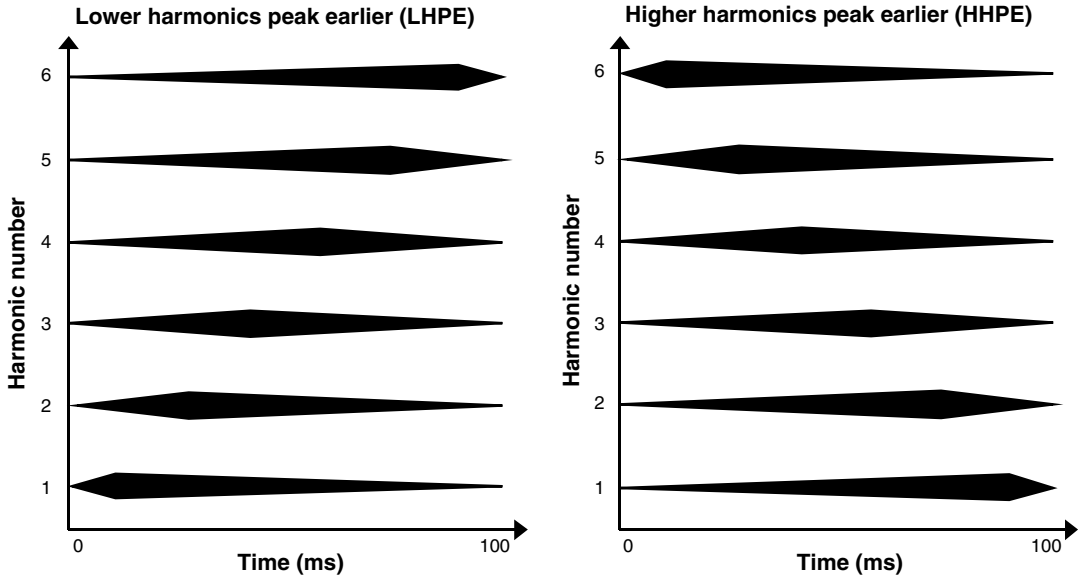

Fig. 1. The amplitude envelopes of the harmonics for the tones used in Experiments 1 and 2 (LHPE and HHPE tones). The thickness of each line indicates the variation in the amplitude of the envelope across time. All harmonics had the same peak amplitude.

individual components are shown in Fig. 1(a). The trajectory of the frequency centroid rises for LHPE tones. In the other pattern of temporal envelopes, the peak time for harmonic n was $10 + 16(6 - n)$ ms, giving 90, 74, 58, 42, 26, and 10 ms for the six harmonics, respectively. This pattern is called "higher harmonics peak earlier" (HHPE), and the amplitude envelopes for the individual components are shown in Fig. 1(b). The trajectory of the frequency centroid falls for HHPE tones. The envelope shapes were chosen so that the total energy in each tone was equal, and so that spectral splatter was kept to a minimum. Note that the HHPE case is equivalent to time reversing the LHPE case. The different harmonics in the sounds had a nearly identical energy content (<0.01 dB variation).

There were four conditions, each with a different combination of the temporal-envelope patterns. These are shown in Table 1. Within each condition, the separation in F_0 between the A and B tones was varied. In all cases, the F_0 of the A tones was 400 Hz. The B tones were presented at one of the three higher F_0s in both the same and different envelope conditions: 2, 3, and 4 semitones above the lower-F_0 tones (449, 476, and 504 Hz, respectively). It is well established that a larger difference in F_0 induces greater streaming (e.g., van Noorden, 1975). The range of F_0 separations used was chosen to optimize the sensitivity of our measure by ensuring that the sequences were typically heard to flip back and forth between one- and two-stream percepts.

Table 1
Conditions used in Experiments 1 and 2

Condition	A tones	B tones
1 (same envelope)	LHPE	LHPE
2 (same envelope)	HHPE	HHPE
3 (different envelope)	LHPE	HHPE
4 (different envelope)	HHPE	LHPE

LHPE, lower harmonics peak earlier; HHPE, higher harmonics peak earlier.

All stimuli were synthesized at a sampling rate of 16 kHz and played back via a 16-bit digital-to-analog converter (Data Translation DT2823). They were low-pass filtered (Fern Developments EF16X module; corner frequency = 5.2 kHz, roll-off = 100 dB/oct) and presented diotically over Sennheiser HD 480-13II earphones. The stimuli had a level of 72 dB SPL (measured at their center in time) and were played to the listeners in a sound-attenuating chamber (Industrial Acoustics).

2.1.4. Procedure

Participants were told that at any particular moment during a trial they were likely to hear the sequence as either a single stream with a galloping rhythm, or as separate high-pitched and low-pitched streams. They were shown a diagram to aid the explanation. Participants pressed a key to begin each trial. They were then asked to respond as soon as they could identify which of the two percepts they heard, using the appropriate response key, and thereafter to respond each time they heard a change. Their current selection was displayed on a screen. Participants were given 12 practice trials comprising one repetition of each combination of 3 F_0s $\times$ 4 temporal-envelope conditions. The main block comprised four repetitions of each stimulus, giving a total of 48 trials, which were presented in random order.

2.2. Results

The first 5 s of each trial were ignored, while listeners initially selected a response. In the trials in which a response had not been made after this period (0.7%), the time until a selection was made was also ignored. The total proportion of the remaining time (quantized in 0.5 s units) that "two streams" was selected was taken as a measure of the amount of stream segregation. The results, averaged over listeners and repetitions, are shown in Fig. 2. In a within-subjects ANOVA using the Huynh–Feldt correction for sphericity, the effects of F_0 separation ($F(2,14) = 46.0$, $p < 0.001$) and of condition ($F(3,21) = 15.6$, $p < 0.001$) were significant. The interaction of these two factors was not significant ($F(6,42) = 1.81$, $p = 0.14$). As expected (e.g., van Noorden, 1975), sequences with a smaller F_0 separation showed less segregation. Bonferroni's protected-t was used to test the a priori prediction regarding the effect of envelope differences and also to make two a posteriori comparisons. First, there was significantly less stream segregation in the conditions where the pattern of spectral variation of the A and B tones was the same (Conditions 1 and 2) than in the conditions where the pattern of spectral variation was different (Conditions 3 and 4) ($t'(7) = 4.57$, $p < 0.01$). Second, there was significantly greater stream segregation in Condition 3 than in Condition 4 ($t'(7) = 5.19$, $p < 0.01$). Third, there was a non-significant trend towards greater segregation in Condition 1 than in Condition 2 ($t'(7) = 2.56$; $p = 0.12$).

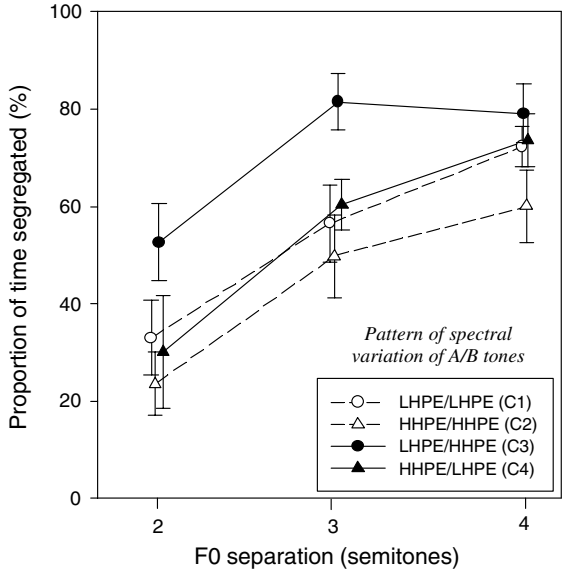

Fig. 2. The results of Experiment 1. The mean proportion of stream segregation (with intersubject standard errors) is shown across eight participants. Open and closed symbols correspond to conditions in which the temporal-envelope pattern for tones A and B was the same or different, respectively. For clarity, the symbols for each condition are shown slightly displaced from one another at each F_0 separation. Condition numbers are also indicated (C1–C4).

2.3. Discussion

The results confirm that a sequence comprising sounds with different patterns of spectral variation over time is more likely to segregate than a sequence comprising sounds with the same pattern of spectral variation over time. Our results are consistent with those of Iverson (1995), who found that the correlation over time of the frequency centroids of sounds in a sequence was a good predictor of their stream segregation and concluded that differences in this factor might enhance streaming.

An alternative hypothesis that might have been proposed is that least stream segregation would be obtained when discontinuities in the frequency centroid between consecutive sounds in a sequence were at a minimum. Fig. 3 shows the frequency centroid over time for ABA triplets from sequences used in Condition 1 (same envelope) and Condition 3 (different envelope), with the largest F_0 separation used (four semitones). It can be seen that the discontinuities of the frequency centroid were larger in Condition 1 than in Condition 3. Condition 2 is just Condition 1 reversed in time and so the size of the discontinuity will have been identical. This symmetry argument also applies to Conditions 3 and 4. From the discontinuity hypothesis, greater segregation would have been expected in the same-envelope conditions (Conditions 1 and 2) than in the different-envelope conditions (Conditions 3 and 4). However, the opposite effect was observed and so this explanation can be rejected. It should be noted that, in this respect, our results parallel those of Steiger and Bregman (1981). They found that sequences of pure-tone glides showed less segregation when the pattern of variation within each sound was similar (e.g., a sequence of upward glides) than when the discontinuity between the end of one glide and the start of the next was minimized (e.g., alternating upward and downward glides).

There is another effect apparent in the results. There was less segregation when the A tones were of the LHPE form (Conditions 1 and 3) than when they were of the HHPE form (Conditions 2 and 4). There are two possible causes of this difference, because there are two asymmetries between the A and B tones. First, owing to the rhythmic pattern used, there were twice as many A tones as B tones. Therefore, one possible factor is that a greater proportion of LHPE tones leads to greater stream segregation. Second, in all conditions the A tones had a lower F_0 than the B tones. Therefore, another possible factor is that the sequence is perceived to have greater stream segregation when the lower-F_0 tones are of the LHPE pattern. To test which of these asymmetries was the cause of the difference in streaming between

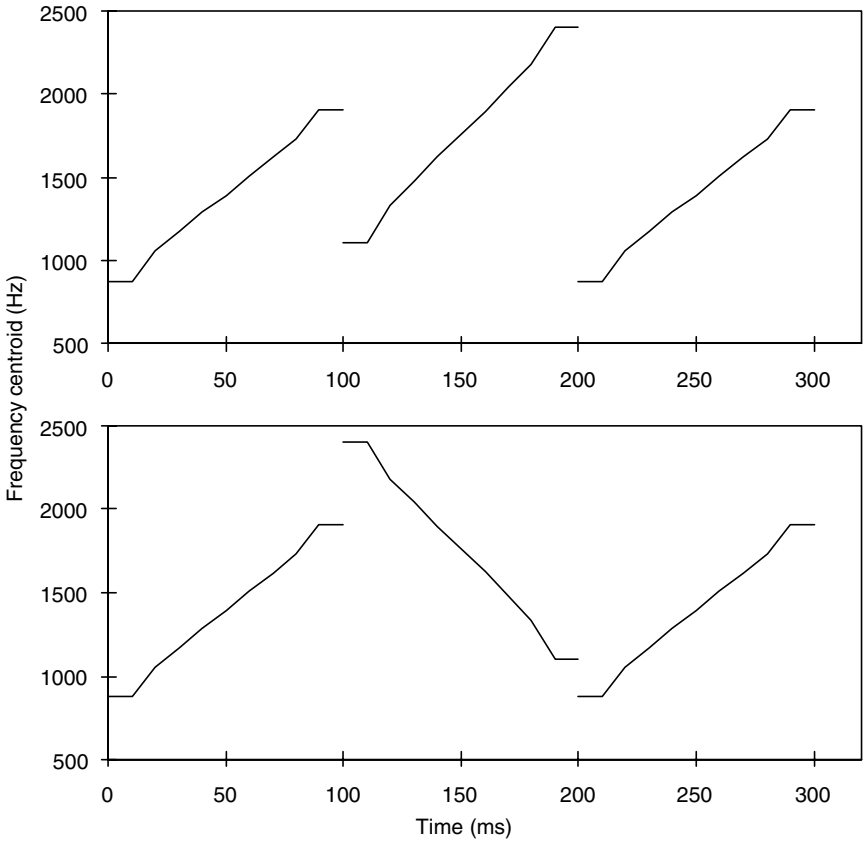

Fig. 3. The variation of the frequency centroid over time for exemplar stimuli used in Experiment 1. The top panel represents an ABA triplet for the LHPE/LHPE case (a same-envelope condition). The bottom panel represents an ABA triplet for the LHPE/HHPE case (a different-envelope condition). The F_0 separation for tones A and B corresponds to the greatest value used (four semitones).

Conditions 1 and 2 and between Conditions 3 and 4, the experiment was repeated but with the A and B tones exchanged in F_0. If the difference were due to the pattern of spectral variation of the more numerous tones, the pattern of results across conditions would be expected to stay the same after this manipulation. However, if the difference were due to the pattern of spectral variation of the lower-F_0 tones, then Conditions 3 and 4 would be expected to exchange places while Conditions 1 and 2 remain unchanged. Other changes in the pattern of results obtained would indicate more complex interactions between the two factors.

Finally, the trend towards greater stream segregation in Condition 1 than in Condition 2 can be related to Iverson's (1995) finding that sounds with more abrupt onsets tend to segregate from each other more than do sounds with gentler onsets. This is because the abrupt attack on the first harmonic for the LHPE tones will be fully resolved from the gentler attacks of its neighbors, whereas the abrupt attack on the sixth harmonic for the HHPE tones will not (see, e.g., Plomp, 1964). Hence, the effective attack may have been gentler for the HHPE tones.

3. Experiment 2

3.1. Method

3.1.1. Participants

Six undergraduate students participated in this experiment, which lasted about 1 h. All reported normal hearing.

3.1.2. Stimuli, apparatus, and procedure

These were the same as for Experiment 1, except that the tones with the higher F_0 were now twice as frequent as the lower tones rather than vice versa. The less numerous tones (the B tones) had an F_0 fixed at 400 Hz and the more numerous A tones were presented at three different F_0s: 2, 3, and 4 semitones above the B tones (449, 476, and 504 Hz, respectively).

3.2. Results

As before, the first 5 s of each trial were ignored, while participants initially selected a response. In the trials in which a response had not been made after this

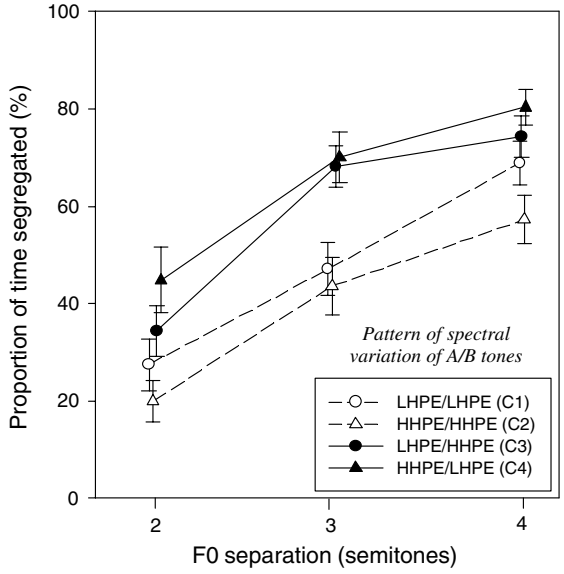

Fig. 4. The results of Experiment 2. The mean proportion of stream segregation (with intersubject standard errors) is shown across six participants. Open and closed symbols correspond to conditions in which the temporal-envelope pattern for tones A and B was the same or different, respectively. For clarity, the symbols for each condition are shown slightly displaced from one another at each F_0 separation. Condition numbers are also indicated (C1–C4).

period (3.5%), the time until a selection was made was also ignored. The results, averaged over listeners and repetitions, are shown in Fig. 4. As in Experiment 1, a within-subjects ANOVA showed a strong effect of F_0 separation ($F(2,10) = 25.6$, $p < 0.001$) and of condition ($F(3,15) = 5.04$, $p < 0.02$). Again, the interaction between these factors was not significant ($F(6,30) = 0.63$). As expected (e.g., van Noorden, 1975), sequences with a smaller F_0 separation showed less segregation. Three comparisons were performed with Bonferroni's protected-t to test the a priori predictions. These are described in relation to the findings for Experiment 1. Once again, there was less stream segregation in the conditions where the pattern of spectral variation of the A and B tones was the same (Conditions 1 and 2) than in the conditions where the pattern of spectral variation was different (Conditions 3 and 4) ($t'(5) = 4.29$, $p < 0.05$). In addition, the non-significant trend towards greater stream segregation in Condition 1 (LHPE tones) than in Condition 2 (HHPE tones) was replicated ($t'(5) = 1.15$). However, the significantly greater stream segregation previously observed for Condition 3 compared with Condition 4 was replaced by a non-significant trend in the opposite direction ($t'(5) = 0.75$).

3.3. Discussion

Once again, similarity of the pattern of spectral variation between tones A and B had a clear effect on streaming. There was less stream segregation when tones in a sequence had the same pattern of spectral variation. Overall, the change across experiments in the relative amount of stream segregation observed for Conditions 3 and 4 suggests that greater stream segregation is induced in the different-envelope conditions when the lower-F_0 tones have the LHPE pattern of spectral variation rather than when tones with the LHPE pattern are more numerous.

An explanation of this effect can be provided in terms of two findings reported in the literature. First, Serafini (1993) observed in her study of the percussive tones used in Javanese gamelan music that the beginning of a sound makes a disproportionately large contribution to its timbre. This effect, which may reflect adaptation in the auditory system, would have caused an LHPE tone to have a lower overall brightness than an HHPE tone with the same F_0, because the early portion of an LHPE tone is dominated by the energy of its low harmonics. Second, there is evidence for an interaction between the effects of differences in spectral center (corresponding to brightness) and in F_0 (corresponding to pitch height) on stream segregation (McAdams and Bregman, 1979; Bregman et al., 1990). The observed trends for Conditions 3 and 4 of Experiments 1 and 2 can be understood by combining these factors. When the sounds with the lower F_0 had the LHPE pattern and those with the higher F_0 the HHPE pattern, the effects of the two factors reinforced each other and greater segregation was observed than when the factors were in opposition.

Experiments 1 and 2 have shown that sounds with similar patterns of dynamic variation undergo less stream segregation than do those with different patterns. However, it is not clear whether this is mediated by changes in the frequency centroid over time, or whether differences in dynamic spectral variation in the absence of changes in the frequency centroid will still enhance segregation. This was investigated in Experiment 3.

4. Experiment 3

4.1. Method

4.1.1. Participants

Eight undergraduate students participated in this experiment, which lasted about 1 h. All reported normal hearing.

4.1.2. Stimuli, apparatus, and procedure

The general structure of the sequences was similar to those used in Experiment 2. All tones comprised the first six harmonics of F_0. Each of the harmonics rose linearly in amplitude over time, reached a peak, and then decreased linearly in amplitude back down to zero. All harmonics began at the same time and were 100 ms long. There were two different patterns of temporal variation for the

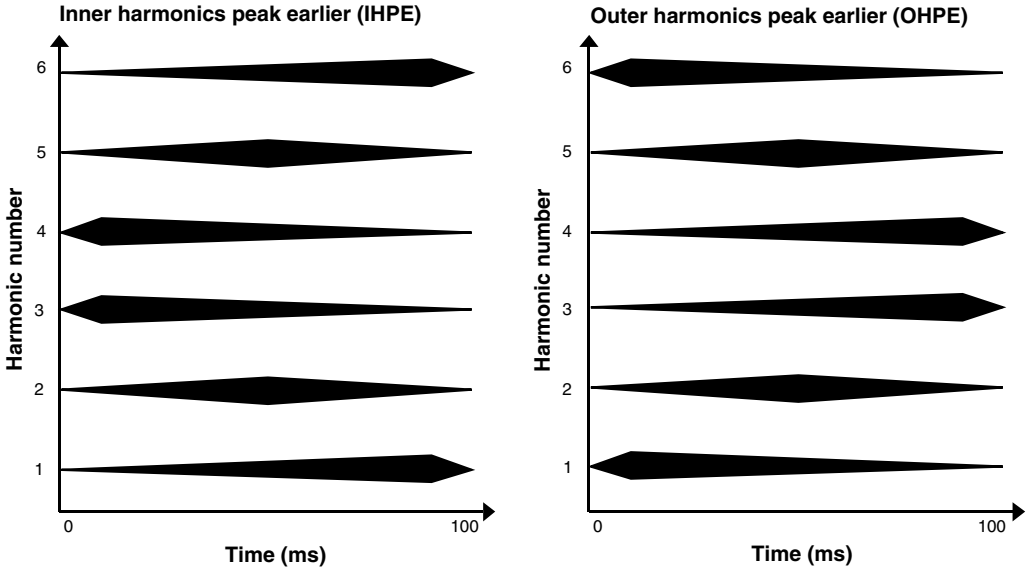

Fig. 5. The amplitude envelopes of the harmonics for the tones used in Experiment 3 (IHPE and OHPE tones). The thickness of each line indicates the variation in the amplitude of the envelope across time. All harmonics had the same peak amplitude.

harmonics of the tones. In one pattern, the peak times were 90, 50, 10, 10, 50, and 90 ms for the six harmonics, respectively. This pattern is called "inner harmonics peak earlier" (IHPE), and the amplitude envelopes for the individual components are shown in Fig. 5(a). In the other pattern of temporal envelopes, the peak times were 10, 50, 90, 90, 50, and 10 ms for harmonics 1, 2, 3, 4, 5, and 6, respectively. This pattern is called "outer harmonics peak earlier" (OHPE), and the amplitude envelopes for the individual components are shown in Fig. 5(b). Again, the envelope shapes were chosen so that spectral splatter was kept to a minimum and so that the total energy in each complex was equal. The different harmonics in the sounds had a nearly identical energy content (<0.01 dB variation). The frequency centroid, as measured using a linear frequency scale, was constant over time and identical for all of the sounds.

There were again four different conditions with different combinations of the temporal-envelope patterns. These are shown in Table 2. As before, within each condition, the separation in F_0 between tones A and B was varied. In all cases, the F_0 of the B tones was 400 Hz. The A tones were presented at three F_0s in each condition: 2, 3, and 4 semitones above the lower tones (449, 476, and 504 Hz, respectively). The apparatus and procedure were the same as for Experiments 1 and 2.

Table 2
Conditions used in Experiment 3

Condition	A tones	B tones
1 (same envelope)	IHPE	IHPE
2 (same envelope)	OHPE	OHPE
3 (different envelope)	IHPE	OHPE
4 (different envelope)	OHPE	IHPE

IHPE, inner harmonics peak earlier; OHPE, outer harmonics peak earlier.

4.2. Results

As before, the first 5 s of each trial were ignored, while participants initially selected a response. In the trials in which a response had not been made after this period (1.5%), the time until a selection was made was also ignored. The results, averaged over listeners and repetitions, are shown in Fig. 6. As for Experiments 1 and 2, a within-subjects ANOVA showed a strong effect of F_0 separation ($F(2,14) = 16.4$, $p < 0.001$) and no interaction of F_0 separation and temporal-envelope condition ($F(6,42) = 0.207$). As expected (e.g., van Noorden, 1975), sequences with a smaller F_0 separation showed less stream segregation. However, in contrast with Experiments 1 and 2, there was no significant effect of temporal-envelope condition ($F(3,21) = 1.43$).

4.3. Discussion

There was no evidence of greater stream segregation when differences in the patterns of spectral evolution of the sounds were not accompanied by differences in the pattern of variation in the frequency centroid over time. This suggests that the effect of envelope differences on stream segregation observed in Experiments 1 and 2 was due to differences in the patterns of variation of the frequency centroid for the sounds rather than to differences in the patterns of spectral variation per se.

It should be noted that measuring the centroid on a logarithmic frequency scale, or an equivalent rectangular bandwidth (ERB) scale (Glasberg and Moore, 1990),

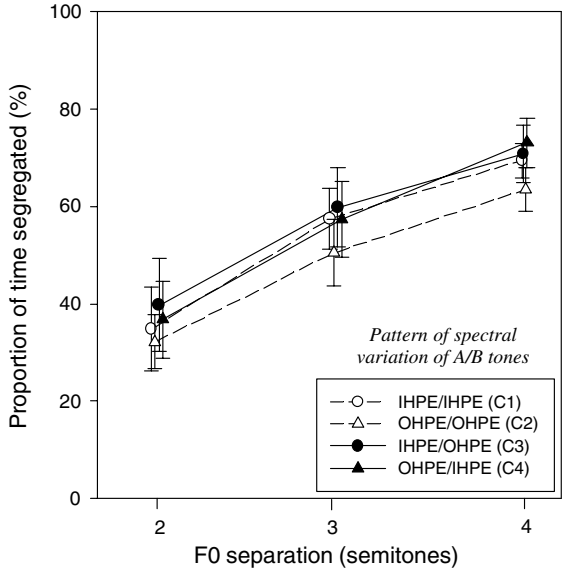

Fig. 6. The results of Experiment 3. The mean proportion of stream segregation (with intersubject standard errors) is shown across eight participants. Open and closed symbols correspond to conditions in which the temporal-envelope pattern for tones A and B was the same or different, respectively. For clarity, the symbols for each condition are shown slightly displaced from one another at each F_0 separation. Condition numbers are also indicated (C1–C4).

might be more appropriate than measuring it on a linear scale. On a log scale, the OHPE and IHPE sounds used in Experiment 3 would have had slightly different patterns of variation in their frequency centroids over time. Specifically, the centroid for an OHPE tone rises from a minimum of $2.6 \times F_0$ (10 ms after onset) to a maximum of $3.3 \times F_0$ (10 ms before offset), whereas an IHPE tone shows the opposite pattern. This difference might account for the small (but not significant) trend towards greater stream segregation observed in the different-envelope conditions.

According to Iverson's (1995) measure of spectral flux, the stimuli used in this experiment had a slightly larger flux than those used in Experiments 1 and 2. As he found greater segregation for sequences where the total amount of flux averaged across all the sounds was larger, we might have expected greater overall segregation in this experiment. However, no such trend was observed as the total mean proportion of segregation collapsed across all conditions was 56%, 53%, and 54% for experiments 1, 2, and 3, respectively. This probably reflects the greater range of spectral flux values across the set of stimuli used by Iverson (1995) compared with our study.

5. General discussion

Differences in the pattern of spectral variation over time between sounds that are played in a sequence can enhance their stream segregation. In particular, an increase in streaming was found when sounds differed in the trajectory of their frequency centroids, in the absence of differences in the total amount of spectral flux (Experiments 1 and 2). No evidence was found for an effect of the pattern of spectral variation in the absence of changes in the frequency centroid (Experiment 3). These results are in agreement with those of Iverson (1995), who found that the amount of correlation between the frequency centroids over time of sequentially presented sounds was a good predictor of their segregation. Unlike Iverson, the experiments reported here used simple and precisely controlled sounds, thus reducing the possibility of the results being confounded by the effects of other acoustic attributes.

Further consideration should be given to whether or not stream segregation is determined by similarity of timbre. Multidimensional scaling studies have shown that sounds with a more similar pattern of spectral variation are more likely to have a similar timbre (Grey, 1977; Grey and Gordon, 1978; Krumhansl et al., see Krumhansl, 1989). Specifically, it was concluded that the important factor was the similarity between sounds in the amplitude envelopes of their harmonics, which is related to the similarity in the total amount of spectral flux. In contrast, Iverson (1995) found that the difference between two sounds in the amount of spectral flux was not a good predictor of the extent to which they would undergo stream segregation. This implies that stream segregation is not determined by timbre per se. However, to draw this conclusion would be premature. It might be that the similarity in the amount of spectral flux between the sounds used in the studies of timbre was confounded with the similarity of variation in their frequency centroids. This could be investigated using a task in which similarity in timbre is measured for simple and precisely controlled sounds, which would allow stimuli to be equated for total amount of spectral flux. Furthermore, Iverson's (1995) claim that similarity between sounds in the total amount of spectral flux is not important for streaming could be confirmed or refuted using a task such as that used in the experiments reported here.

Our results fit in with a growing picture in the literature, which suggests that perceptual organization is affected by representations at many different levels in the auditory processing pathways. Many streaming effects can be explained by differences in the stimuli that are extracted in the cochlea (Hartmann and Johnson, 1991; Beauvois and Meddis, 1991). However, information not extracted until higher levels can also affect streaming, such as the pitch evoked by unresolved harmonic complexes (Vliegen et al., 1999; Roberts et al., 2002) or by amplitude-modulated broadband noise bursts (Grimault et al., 2002). Even stored representations, such as those of speech sounds, may influence auditory grouping (Scheffers, 1983). Our results fit in with these findings, showing that those parts of the auditory system that are

responsible for extracting timbral attributes, perhaps the thalamus or early cortical areas, can also affect grouping. Perhaps the most attractive model of perceptual organization is one where competition for grouping happens at multiple neural levels on the basis of different kinds of information and then these interact in some way (perhaps by synchronization of neural firing; e.g., Wang, 1996) to generate the final organization.

To summarize, sounds in a sequence that differed in their patterns of spectral variation (but not in the total amount of spectral flux) were found to segregate more than those with the same patterns of spectral variation (Experiments 1 and 2). In these experiments, the sounds differed in the pattern of change of their frequency centroids over time. In contrast, sequences of sounds with different patterns of spectral variation, but with similar and unchanging frequency centroids, were not found to stream any more than those with the same pattern of spectral variation (Experiment 3). In conclusion, differences in the patterns of spectral variation of sounds in a sequence can enhance their streaming. However, differences in the patterns of variation of their frequency centroids over time may well be a necessary condition for this effect on perceptual grouping.

Acknowledgements

We thank Peter Bailey, Al Bregman, Ian Cross, Steve McAdams, and Brian Moore for their comments and suggestions regarding this work.

References

Anstis, S., Saida, S., 1985. Adaptation to auditory streaming of frequency-modulated tones. J. Exp. Psychol. Hum. Percept. Perform. 11, 257–271.

Beauvois, M.W., Meddis, R., 1991. A computer model of auditory stream segregation. Quart. J. Exp. Psychol. 43A, 517–541.

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

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

Bregman, A.S., Liao, C., Levitan, R., 1990. Auditory grouping based on fundamental frequency and formant peak frequency. Can. J. Psychol. 44, 400–413.

Carlyon, R.P., Cusack, R., Foxton, J.M., Robertson, I.H., 2001. Effects of attention and unilateral neglect on auditory stream segregation. J. Exp. Psychol. Hum. Percept. Perform. 27, 115–127.

Cusack, R., 1998. The Role of Differences in the Temporal Characteristics of Sounds on their Sequential Grouping. Doctoral thesis, University of Birmingham, Birmingham, UK.

Cusack, R., Roberts, B., 1999. Effects of similarity in bandwidth on the auditory sequential streaming of two-tone complexes. Perception 28, 1281–1289.

Cusack, R., Roberts, B., 2000. Effects of differences in timbre on sequential grouping. Percept. Psychophys. 62, 1112–1120.

Glasberg, B.R., Moore, B.C.J., 1990. Derivation of auditory filter shapes from notched-noise data. Hear. Res. 47, 103–138.

Grey, J.M., 1977. Multidimensional perceptual scaling of musical timbres. J. Acoust. Soc. Am. 61, 1270–1277.

Grey, J.M., Gordon, J.W., 1978. Perceptual effects of spectral modifications on musical timbres. J. Acoust. Soc. Am. 63, 1493–1500.

Grimault, N., Bacon, S.P., Micheyl, C., 2002. Auditory stream segregation on the basis of amplitude-modulation rate. J. Acoust. Soc. Am. 111, 1340–1348.

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

Hartmann, W.M., Johnson, D., 1991. Stream segregation and peripheral channeling. Music Percept. 9, 155–183.

Iverson, P., 1995. Auditory stream segregation by musical timbre: effects of static and dynamic acoustic attributes. J. Exp. Psychol. Hum. Percept. Perform. 21, 751–763.

Iverson, P., Krumhansl, C.L., 1993. Isolating the dynamic attributes of musical timbre. J. Acoust. Soc. Am. 94, 2595–2603.

Krumhansl, C.L., 1989. Why is musical timbre so hard to understand? In: Nielzen, S., Olsson, O. (Eds.), Structure and Perception of Electroacoustic Sound and Music. Elsevier, Amsterdam.

Kruskal, J.B., 1964a. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29, 1–27.

Kruskal, J.B., 1964b. Nonmetric multidimensional scaling: a numerical method. Psychometrika 29, 115–129.

McAdams, S., Bregman, A.S., 1979. Hearing musical streams. Comp. Music J. 3, 26–43.

Moore, B.C.J., Gockel, H., 2002. Factors influencing sequential stream segregation. Acustica – Acta Acustica 88, 320–333.

Plomp, R., 1964. The ear as a frequency analyzer. J. Acoust. Soc. Am. 36, 1628–1636.

Risset, J.-C., Wessel, D.L., 1999. Exploration of timbre by analysis and synthesis. In: Deutsch, D. (Ed.), The Psychology of Music, second ed. Academic Press, San Diego.

Roberts, B., Glasberg, B.R., Moore, B.C.J., 2002. Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. J. Acoust. Soc. Am. 112, 2074–2085.

Scheffers, M.T.M., 1983. Sifting vowels: auditory pitch analysis and sound segregation. Doctoral thesis, Groningen University, The Netherlands.

Serafini, S., 1993. Timbre perception of cultural insiders: a case study with Javanese gamelan instruments. Masters thesis, University of British Columbia, Canada.

Shepard, R.N., 1962a. The analysis of proximities: multidimensional scaling with an unknown distance function I. Psychometrika 27, 125–139.

Shepard, R.N., 1962b. The analysis of proximities: multidimensional scaling with an unknown distance function. II. Psychometrika 27, 219–246.

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., Bregman, A.S., 1997. The influence of different timbre attributes on the perceptual segregation of complex-tone sequences. J. Acoust. Soc. Am. 102, 1943–1952.

Steiger, H., Bregman, A.S., 1981. Capturing frequency components of glided tones: frequency separation, orientation, and alignment. Percept. Psychophys. 30, 425–435.

van Noorden, L.P.A.S., 1975. Temporal coherence in the perception of tone sequences. Doctoral thesis, Eindhoven University of Technology, The Netherlands.

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. J. Acoust. Soc. Am. 106, 938–945.

Wang, D.L., 1996. Primitive auditory segregation based on oscillatory correlation. Cognit. Sci. 20, 409–456.

Wessel, D.L., 1979. Timbre space as a music control structure. Comp. Music J. 3, 45–52.