Review

Properties of auditory stream formation

Brian C. J. Moore[1],* and Hedwig E. Gockel[2]

[1]*Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK*
[2]*MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, UK*

A sequence of sounds may be heard as coming from a single source (called fusion or coherence) or from two or more sources (called fission or stream segregation). Each perceived source is called a 'stream'. When the differences between successive sounds are very large, fission nearly always occurs, whereas when the differences are very small, fusion nearly always occurs. When the differences are intermediate in size, the percept often 'flips' between one stream and multiple streams, a property called 'bistability'. The flips do not generally occur regularly in time. The tendency to hear two streams builds up over time, but can be partially or completely reset by a sudden change in the properties of the sequence or by switches in attention. Stream formation depends partly on the extent to which successive sounds excite different 'channels' in the peripheral auditory system. However, other factors can play a strong role; multiple streams may be heard when successive sounds are presented to the same ear and have essentially identical excitation patterns in the cochlea. Differences between successive sounds in temporal envelope, fundamental frequency, phase spectrum and lateralization can all induce a percept of multiple streams. Regularities in the temporal pattern of elements within a stream can help in stabilizing that stream.

Keywords: perceptual stream formation; sound sequences; fusion; fission; bistability

1. INTRODUCTION: THE PHENOMENOLOGY OF STREAMING AND BISTABILITY

A source is a physical entity that gives rise to sound, for example, a violin being played. A stream is the percept of a group of successive and/or simultaneous sound elements as a coherent whole, appearing to emanate from a single source. For example, it is the percept of hearing a violin being played [1]. When listening to a rapid sequence of sounds, the sounds may be perceived as a single stream (called fusion or coherence), or they may be perceived as more than one stream (called fission or stream segregation) [1,2]. Here, the term 'streaming' is used to denote the processes that determine whether one stream or multiple streams are heard. This paper reviews psychoacoustic studies of the properties of stream formation and the factors that influence streaming.

Streaming depends on the amount of difference between successive sounds and the rate of presentation of the sounds. In many studies, the sounds have been sinusoidal tones, and fission has been induced by differences in frequency between successive tones, but other differences (e.g. in temporal envelope) between sounds can also induce fission. For the moment, we consider the effect of frequency separation as an example. Large frequency separations and high presentation rates tend to lead to fission, while small separations and low rates tend to lead to fusion. These effects may indicate that very rapid and large changes in sounds are interpreted as evidence for a change in sound source and lead to the perception of fission, while slow and small changes are interpreted as changes within a single source and lead to the perception of a single stream.

For intermediate differences between successive sounds in a rapid sequence, either fission or fusion may be heard. This has often been examined using tone sequences of the form ABA−ABA−..., where A and B represent brief sinusoidal tone bursts and − represents a silent interval [2]. Figure 1 illustrates such a sequence. When the frequencies of A and B are very different, as at the start of the sequence, two streams are usually heard (fission), one (A tones) going twice as fast as the other (B tones). When the frequencies of A and B are similar, fusion usually occurs and a characteristic 'gallop' rhythm is heard. For intermediate frequency separations of A and B, in the 'ambiguity region', the percept may often 'flip' between one and two streams. This is called 'bistability'. The flips do not occur in a regular manner [3]. Listeners usually report that they do not hear both percepts at the same time, although Bendixen *et al.* [4] found that listeners did sometimes report hearing both an integrated stream consisting of high and low tones and a separate stream consisting of only high or only low tones. To some extent, perception in the ambiguity region can be influenced by the instructions to the listener. For example, the listener may be instructed to try to hear a single stream or to try to hear two streams, and subjective reports indicate that this can influence the relative proportion of time for which fusion and fission are heard [2].

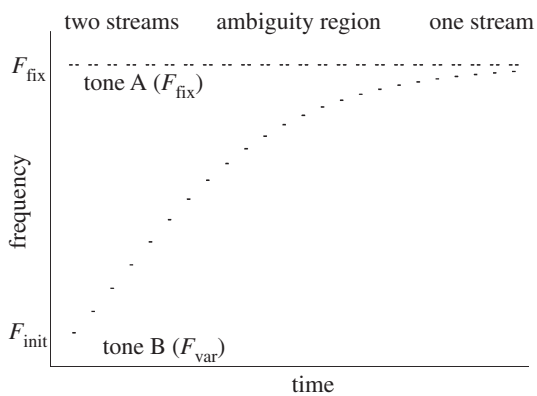* Author for correspondence (bcjm@cam.ac.uk).

Figure 1. Illustration of how the percept of a tone sequence is affected by the frequency separation of the tones. The sequence is of the form ABA–ABA–. In this example, the frequency of tone A is fixed at F_{fix} and the frequency of tone B starts well below that of tone A, at frequency F_{init}, and is moved towards that of tone A. When the frequency separation of A and B is large, two streams are usually heard. When the frequency separation is small, one stream is usually heard.

2. TASKS USED TO MEASURE STREAMING

Some experiments examining streaming have required only subjective responses from the listeners. For example, when listening to the ABA–ABA . . . sequence, listeners may be asked to indicate whether or not they hear the gallop rhythm. However, some researchers have used a more objective approach in which performance is examined in a task that is assumed to be affected by streaming. For some tasks, performance should be better if fission occurs. For example, a task requiring the recognition of two interleaved melodies [5,6] is performed best if the tones making up the two melodies are heard as separate streams. Thus, the subject may try to hear segregation to improve performance. In contrast, some tasks may be performed more poorly if segregation occurs. For example, it is difficult to compare the timing of sound elements in different streams [7–9]. Thus, a task requiring judgements of the timing or rhythm of sounds in a sequence may be more difficult if the sounds to be compared are heard as part of separate streams than if they are heard as a single stream [7,10–12]. Fission that occurs even when the listener is trying to hear fusion is sometimes called 'primitive stream segregation' [1,10,11] or 'obligatory segregation' [12]. Tasks that are performed more poorly when segregation is heard are used to measure obligatory segregation.

3. PRIMITIVE VERSUS SCHEMA-BASED SEGREGATION

Bregman [1] distinguished between primitive and schema-based mechanisms in stream segregation. He thought of the former as a bottom-up, pre-attentive, sensory partitioning mechanism for which the auditory system uses basic cues that are present in the stimulus, for example whether the spectrum is similar from one sound to the next or changes markedly across sounds. He thought of the latter as a top-down, selection

mechanism for which acquired knowledge about sound is used to process and organize the incoming sensory information.

Bey & McAdams [13] investigated the role of schema-based processes in streaming. They used a melody recognition task in which two unfamiliar six-tone sequences (the target melodies) were presented successively to listeners. One of these was interleaved with distractor tones (hereafter called the 'mixture'). Listeners had to indicate whether the two target melodies were the same or different. In one condition, the target alone was presented first; thus, listeners had precise knowledge about the target melody when the mixture was presented. In another condition ('post-recognition condition'), the mixture was presented first; thus, listeners did not know in advance the target melody that they had to try to hear out from the mixture. Bey & McAdams measured performance as a function of the mean frequency difference between the target and the distractor tones. As expected, performance improved with increasing frequency difference. Performance was much better when the target alone was played first than when it was played second, but only when there was a difference in mean frequency between the target and the distractor tones. This was interpreted as evidence for a schema-driven effect that can operate only when some primitive segregation has occurred. However, this interpretation was questioned by Devergie *et al.* [14]. They used an interleaved-melody task with highly familiar melodies and showed that performance was above chance when the frequencies of the target and distractor tones fell in the same range. It appears that a schema-driven effect can operate without primitive segregation when the melodies are highly familiar, but not when they are unfamiliar and are cued only by a preceding tone sequence.

4. THE BUILD-UP, RESETTING AND DECAY OF STREAM SEGREGATION
(a) *The build-up of stream segregation*
When a long sound sequence is presented with an intermediate difference between the sounds in the sequence, the tendency for fission to occur increases with increasing exposure time to the tone sequence [15,16]. One interpretation of this is that the auditory system starts with the assumption that there is a single sound source, and fission is perceived only when sufficient evidence has built up to contradict this assumption. For the tone sequences typically used in experiments on streaming (with alternation rates in the range 2–10 per second), the tendency to hear fission increases rapidly at first, but slows after about 10 s [16]. If the frequency separation between tones in the sequence is larger than a certain value, called the temporal coherence boundary (TCB [2]), two streams are usually heard, at least after a few tones have been presented; this happens even when the subject is instructed to try to hear a single stream. If the frequency separation is less than a (smaller) critical value, called the fission boundary (FB [2]), a single stream is usually heard, even when the subject is instructed to try to hear two streams. However, even for very small or very large frequency separations,

the percept may flip if the sequence is presented for a long time [3,17]. In other words, bistability occurs. Thus, like most 'thresholds', the TCB and FB are not 'hard' boundaries; they are defined in terms of the probability of hearing fusion or fission.

The FB is affected only slightly by the repetition period of the tones, while the TCB increases markedly with increasing repetition period [2]. Bregman *et al.* [18] showed that the most important temporal factor influencing the TCB was the time interval between successive tones of the same frequency (e.g. between the A tones in the sequence ABA–ABA), rather than the interval between tones of different frequency or the duration of the tones.

(b) *Factors influencing the build-up, resetting and decay of the tendency for segregation*

Several factors can influence the tendency to hear stream segregation at a given point during a sequence. Anstis & Saida [16] presented sequences of tones with alternating frequency (ABAB...) to one ear, using sequence lengths sufficient to allow segregation to build up. They then switched the sequence suddenly to the other ear. The sequence was perceived as one stream following the switch. The tendency to perceive fission seemed to be 're-set' by the sudden change in the ear of presentation.

The factors governing this 're-setting' were explored by Rogers & Bregman [19]. In one of their experiments, they presented a sequence of tones composed of a relatively long induction sequence (intended to allow stream segregation to build up) and a shorter test sequence, with no pause between them. The test sequence was similar to the ABA–ABA sequence described above; it was presented monaurally to the right ear only. The characteristics of the induction sequence were varied. The subjects were instructed to try to hear a single stream ('listen for a gallop') and to rate the strength of segregation at the end of the test sequence. The ratings were highest when the induction sequence was identical to the test sequence and was presented monaurally to the same ear as the test sequence. When the induction sequence was presented diotically (the same sound to each ear), producing a shift in the perceived location and an increase in loudness of the induction sequence (while the test sequence remained monaural), the segregation rating was markedly reduced. Thus, although the right ear received the same input as in the 'identical' condition, the shift in perceived location or loudness produced by adding the induction tones to the contralateral ear led to reduced segregation.

Rogers & Bregman [20] also used a relatively long induction sequence and a shorter ABA–ABA test sequence. They measured the TCB at the end of the test sequence. The TCB was found to be smallest (the build-up of segregation was strongest) when the induction and test sequences were identical. When the induction and test sequences differed in perceived location (for example, because the sounds came from different loudspeakers), TCBs were markedly increased. When the induction sequence was perceived at one location and the test sequence at another, the build-up of segregation was increased if the transition from one location to the other was gradual rather than abrupt. The TCB was also increased (segregation was reduced) when the intensity of the test sequence was suddenly increased relative to that of the induction sequence; however, this did not occur when the intensity was suddenly reduced.

Roberts *et al.* [21] studied the build-up and resetting of streaming using an objective rhythm-discrimination task, for which performance was expected to be relatively poor when obligatory segregation occurred. A 2.0 s fixed-frequency inducer sequence was followed by a 0.6 s test sequence of alternating pure tones (3 low (L)–high (H) cycles). Listeners compared intervals for which the test sequence was either isochronous (had a regular rhythm) or the H tones were slightly delayed (making the rhythm anisochronous or irregular) and were asked to identify the interval in which the H tones were delayed. Resetting of segregation should make identifying the anisochronous interval easier. The H–L frequency separation was varied (0–12 semitones) and the properties of the inducer sequence were varied. The inducer properties manipulated were frequency (the same as for the L tones or two octaves lower), number of onsets (several short bursts versus one continuous tone), the tone/silence ratio (short versus extended bursts), level and lateralization (apparent position in space, manipulated by varying the interaural time difference (ITD)—the relative time of arrival at the two ears). All differences between the inducer and the L tones in the test sequence reduced temporal discrimination thresholds towards those for the no-inducer case, including properties shown previously not to affect subjective ratings of segregation greatly. In other words, differences between the inducer and the L tones led to improved performance, indicating reduced segregation of the tones in the test sequence. Roberts *et al.* concluded that a large variety of types of abrupt changes in a sequence can cause resetting and improve subsequent temporal discrimination.

One interpretation of these results is that sudden changes in a sequence indicate the activation of a new sound source and this causes the percept to revert to its initial 'default' condition, which is fusion. Consistent with this idea, Haywood & Roberts [22] showed that a single deviant tone (a change in frequency, duration or replacement with silence) at the end of an induction sequence can lead to substantial resetting of build-up. They proposed that a single change actively resets the build-up evoked by the induction sequence. However, it is clear that resetting is not an all-or-none phenomenon, as it tended to increase with increasing frequency change of the deviant tone.

Beauvois & Meddis [23] investigated how the built-up tendency towards segregation decays over time. Subjects listened to a 10 s induction sequence of repeated tones (AAAA...) designed to build up a tendency towards hearing an A stream. The induction sequence was followed immediately by a silent interval (0–8 s), and then a short ABAB... test sequence. Subjects were asked to indicate whether or not the test sequence was heard as separate A and B streams. As the silent interval between the induction and test sequence was increased, the tendency to hear segregation decreased.

Somewhat different results were obtained by Cusack *et al.* [24]. They presented thirty-three 10 s sequences of sounds with an ABA–ABA pattern, separated by gaps with durations of 1, 2, 5 or 10 s, selected in random order. Subjects indicated whether they heard one stream or two at various times during each sequence. Data for the first sequence were discarded. At the start of each subsequent sequence, one stream was usually reported regardless of the duration of the gaps between sequences. Cusack *et al.* argued that the build-up of stream segregation was reset by even a short gap. However, during the last half of each sequence, the number of two-stream percepts reported was consistently higher when the preceding gap was 1 s than when it was 10 s. This suggests that there was a decaying influence of the preceding sequence.

Evidence for a persisting effect of previous stimuli was also provided by Snyder *et al.* [25]. Subjects listened to an ABA–ABA sequence lasting 10.8 s, with a difference in frequency, Δf, between A and B, and were asked to indicate whether they heard two streams by the end of the sequence. Each sequence began 2 s after the subject made a response for the previous sequence. Larger values of Δf in preceding sequences led to less segregation for the current sequence, for values of Δf in the ambiguity region. The effect was smaller, but still present, when the interval between sequences was increased to 5.76 s. A series of supplementary experiments suggested that the effect was not solely due to response bias. The authors concluded that the context effect lasts for tens of seconds, although it starts decaying after just a few seconds.

In summary, in the absence of sudden changes in the properties of a sequence, the tendency to hear stream segregation builds up rapidly over about 10 s and then continues to build up more slowly up to at least 60 s. It is as if evidence in favour of there being more than one source is built up over time. Changes in a sound sequence may cause a re-setting of the build-up of stream segregation, perhaps because the changes are interpreted as activation of a new sound source. A gap in a sequence may lead to a partial resetting of the build-up of streaming, but it may take many seconds of silence for the effect to decay completely.

(c) The role of attention in the build-up of the tendency for segregation

There is an ongoing debate about whether attention to a sequence of sounds is necessary for a build-up of the tendency to hear segregation. Carlyon *et al.* [26] investigated the role of attention in the build-up of segregation using tone sequences of the type ABA–ABA. The sequences were presented to the left ear of listeners for 21 s. In a baseline condition, no sounds were presented to the right ear. Listeners indicated when they heard a galloping rhythm (one stream) and when they heard two separate streams. In the main experimental condition ('two-task condition'), subjects were required to make judgements about changes in the amplitude of noise bursts presented to the right ear for the first 10 s of the sequence. After 10 s, listeners had to switch task,

and judge the tone sequence in the left ear. In a control condition, the same stimuli were presented as in the two-task condition, but listeners were instructed to ignore the noise bursts in the right ear and just concentrate on the task and stimuli in the left ear.

For the baseline and the control conditions, results were similar and showed the usual build-up of segregation. However, in the two-task condition *after* switching attention from the right to the left ear, the probability of hearing two streams was significantly smaller than when listeners paid attention to the tone sequence in the left ear over the whole time of presentation. In a later study using a similar paradigm, Carlyon *et al.* [27] again presented a long tone sequence, but initially directed attention away from the sequence using a visual task or by asking subjects to count backwards in threes, a task that imposes a substantial cognitive load. After 10 s, subjects were asked to judge whether the tone sequence was heard as one or two streams. Carlyon *et al.* found that the stream segregation was lower than when there was no distracting task during the initial 10 s.

Carlyon *et al.* [26] interpreted their findings as evidence that the build-up of stream segregation depends on the listener paying attention to the tone sequence. However, as pointed out by Moore & Gockel [28], the act of switching attention to the tone sequence may cause a re-setting of the build-up of stream segregation; the effect could be similar to what occurs when the characteristics of the sequence are changed suddenly, as described earlier. Indeed, using a task similar to that of Carlyon *et al.* [26], Cusack *et al.* [24] reported a re-setting of the build-up of stream segregation if attention was briefly diverted away from the ABA–ABA sequence presented to the left ear. As discussed by Cusack *et al.* [24], it is not clear whether the re-setting occurred when attention was diverted away from the ABA–ABA sequence or when it was directed back to it from the different task.

Recently, effects of switching attention between stimuli have been demonstrated using an objective task. Thompson *et al.* [29] required listeners to detect a short delay of a B tone, within one ABA triplet, in an otherwise regular 13.5 s ABA–ABA sequence. The ABA triplet with the delayed B tone could occur either early (at 2.5 s) or late (at 12.5 s) within the sequence. The delay was harder to detect when it was late than when it was early, owing to the build-up of stream segregation. Initially, the ABA–ABA sequence was presented to one ear, and noise bursts were presented to the other ear. The noise bursts stopped after 10 s. Listeners either had to attend to the ABA–ABA sequence throughout (and perform the delay-detection task), or they had to attend to the other ear for the first 10 s and perform a discrimination task on the noise. They then switched attention to the ABA–ABA sequence and performed the delay-detection task. In the attention-switch condition, detection of a delay of the B tone that occurred late during the sequence was much better than detection of late-occurring delays when attention was focused on the ABA–ABA sequence throughout. Indeed, performance was as good as for an early delay when attention was focused on the ABA–ABA

sequence throughout, although the latter comparison could be affected by the fact that there was noise present at the early but not at the late occurrence time.

Overall, the findings of Carlyon *et al.* [26,27], Cusack *et al.* [24] and Thompson *et al.* [29] show that the build-up of segregation can be reduced either by the absence of attention or by a switch of attention, or a combination of the two.

Evidence that some stream segregation can occur in the absence of (full) attention to the auditory sequence comes from a paradigm called the 'irrelevant sound effect' [30]; a task-irrelevant background sound can interfere with serial recall for visually presented items. The amount of interference from the unattended sound depends on its characteristics. Repetition of a single sound disrupts performance less than a sequence of sounds in which each sound, if attended, would be perceived as a discrete entity differing from the one preceding it. Jones *et al.* [31] and Macken *et al.* [32] used ABA–ABA sequences as task-irrelevant background sounds. They varied either the frequency separation between the A and B tones or their presentation rate. They found that task-irrelevant ABA–ABA sequences that are usually perceived as two streams (each of which would be heard as a sequence of identical elements) impaired serial recall less than sequences that are usually perceived as one stream (for which successive elements would be heard as different). This was taken to indicate that some primitive organization into two streams can occur when the sound sequence is outside the focus of the listener's attention.

5. THE ROLE OF PERIPHERAL CHANNELLING IN STREAM SEGREGATION

The cochlea contains multiple bandpass filters with overlapping passbands. These perform a running spectral analysis of sounds. At any given time, the magnitude of the output of the filters, specified as a function of the centre frequency of the filters, is called the excitation pattern [33,34]. It has been proposed [6,35,36] that a large degree of overlap of the excitation patterns evoked by successive sounds tends to produce fusion while a small degree of overlap tends to produce fission.

For sequences of pure tones, the frequency separation of successive tones has a strong influence on streaming, consistent with the idea that overlap of excitation patterns plays a strong role. Rapid sequences of pure-tone-like percepts evoked by applying brief changes (in amplitude or ITD) to individual components of a complex tone are also organized into perceptual streams based on frequency proximity [37].

It is possible that similarity in the pitch of successive tones is the critical factor; for pure tones, frequency and pitch are inextricably linked. However, the results of several early studies suggest that spectral similarity rather than similarity in pitch is the more important factor. A complex tone with fundamental frequency F0 has the same pitch as a pure tone with frequency equal to F0, but has different spectral content to the pure tone. Van Noorden [2] found that listeners always perceived two streams if a pure tone was alternated with a complex tone with F0

equal to the frequency of the pure tone, or if two complex tones with harmonics in different frequency regions but with the same F0 were alternated. He concluded that contiguity 'at the level of the cochlear hair cells' was a necessary (but not sufficient) condition for fusion to occur. Hartmann & Johnson [6] studied streaming using interleaved melodies. Several conditions were tested, using sequences in which successive sounds differed in temporal envelope, spectral composition, ITD and ear of presentation. They found that conditions for which the most peripheral 'channelling' would be expected (i.e. conditions where successive tones differed in spectrum or were presented to opposite ears) led to the best performance and concluded that 'peripheral channelling is of paramount importance' in determining streaming.

Peripheral frequency analysis is often characterized in terms of the equivalent rectangular bandwidth (ERB) of the auditory filters. The average value of the ERB for young listeners with normal hearing is denoted as ERB_N [34]. The value of ERB_N increases with increasing centre frequency, and for centre frequencies above about 1000 Hz has a value of 12–15% of the centre frequency. A frequency scale related to ERB_N can be derived by using the value of ERB_N as the unit of frequency. For example, the value of ERB_N for a centre frequency of 1 kHz is about 130 Hz, so an increase in frequency from 935 to 1065 Hz represents a step of one ERB_N. The scale derived in this way is called the ERB_N-number scale [34]. For brevity, ERB_N number is denoted by the unit 'Cam', following a suggestion of Hartmann [38]. For example, a frequency of 1000 Hz corresponds to 15.59 Cams.

The model of Beauvois & Meddis [35], which is based on the concept of peripheral channelling, predicts that, for a rapid sequence of sinusoids, the FB should correspond to a constant difference on the Cam scale. To test this idea, Rose & Moore [39] used a tone sequence of the form ABA–ABA. Tone A was fixed in frequency at 62, 125, 250, 500, 1000, 2000, 4000, 6000 or 8000 Hz. Tone B started with a frequency well above that of tone A, and its frequency was swept towards that of tone A so that the frequency separation between them decreased in an exponential manner. Subjects were asked to try to hear two streams and to indicate when they heard only a single stream; this point corresponds to the FB.

The results are shown in figure 2. The FB was roughly independent of the frequency of tone A when expressed as the difference in Cams between A and B, which is consistent with the theory of Beauvois & Meddis [35]; the FB corresponded to a change of about 0.4–0.5 Cams. In fact, the FB varied much less across frequency when expressed in Cams than when expressed in Barks (a unit conceptually similar to the Cam, but based on different data, [40]) or in cents (one-hundredths of a semitone).

The influence of the overlap of excitation patterns on streaming was investigated by Bregman *et al.* [41]. They presented alternating band-pass-filtered noise bursts with sharp band edges. The successive bursts differed in centre frequency, but were matched in overall level and in relative bandwidth (bandwidth divided by centre frequency). Listeners were asked to
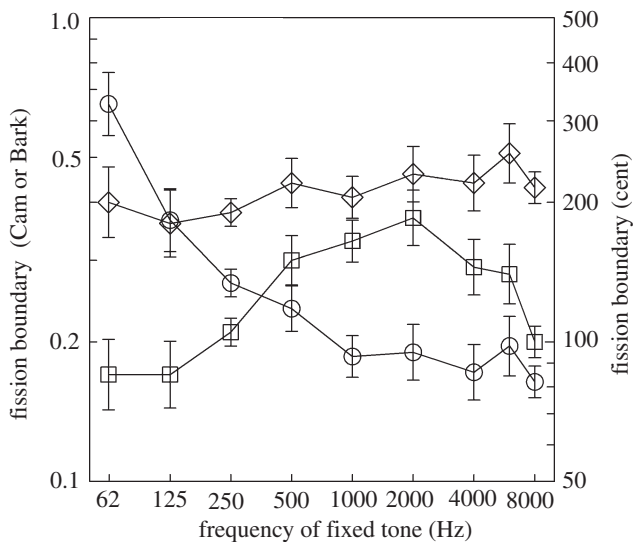
Figure 2. Results of Rose & Moore [39] showing the fission boundary expressed as difference in the number of Cams or Barks (diamonds and squares, respectively, referred to the left axis) or in cents (circles, referred to the right axis) between the A and B tones, and plotted as a function of the frequency of the fixed tone.

judge the ease with which the sequence could be heard as a single stream. Perceived segregation was affected only slightly by the width of the bands, even when the bandwidth was sufficiently large that there was substantial spectral overlap between successive bands. As expected, perceived segregation increased markedly with increasing difference in centre frequency of the two bands. These results indicate that although differences in the excitation patterns of successive sounds can promote stream segregation, overlap of excitation patterns does not prevent segregation.

Cusack & Roberts [42] investigated the effect of dynamic variations in the spectrum of complex tones (comprising harmonics 1–6) created by applying different amplitude envelopes to each harmonic. Listeners judged continuously the grouping of long sequences of ABA–ABA– tones. Tones A and B always differed in F0, and had either the same or different patterns of spectral variation. Sequences with different patterns were more likely to be perceived as segregated than sequences with the same patterns when this resulted in different patterns of change in the spectral centroid over time. However, when the pattern of the spectral centroid was unaltered, segregation was not affected by whether the pattern of spectral variation was the same across the A and B tones. This finding suggests that changes in spectral centroid increase the tendency for segregation, but dynamic spectral changes in spectrum *per se* do not have a strong effect.

6. THE ROLE OF FACTORS OTHER THAN PERIPHERAL CHANNELLING

The evidence reviewed above suggests that peripheral channelling does play a strong role in streaming, but other factors also play a role. Further evidence for these other factors is reviewed below.

(a) *Segregation based on temporal envelope*

Several studies have shown that streaming can be influenced by differences in the temporal envelopes of successive sounds. In some studies, the envelope has been manipulated by varying the bandwidth of a noise or by using both a narrowband noise and a sinusoid as stimuli. A sinusoid has a 'flat' envelope, apart from the onset and offset ramps. A noise has random envelope fluctuations, and the fluctuations increase in rate with increasing bandwidth.

Dannenbring & Bregman [43] alternated two sounds, A and B. These could be both pure tones, both narrowband noises (bandwidth = 9% of the centre frequency), or a tone and a narrowband noise. The excitation pattern of a noise is very similar to that of a tone when the bandwidth is less than 1 ERB_N [44], so very little peripheral channelling would be expected for a noise alternating with a tone of the same centre frequency. Listeners were asked to rate the degree of stream segregation of the sounds. Segregation increased as the frequency separation of A and B was increased. Importantly, segregation was also greater for tone–noise combinations than for tone–tone or noise–noise combinations. Presumably this happened because the envelope of the noise was different from that of the tone, leading to a difference in sound quality (timbre). For sounds with the same centre frequency (1000 Hz) but different timbre, the degree of stream segregation was similar to that for sounds of the same timbre but with moderate frequency differences, equivalent to differences of about 1 Cam. This is larger than the FB for sequences of pure tones, which is 0.4–0.5 Cams [39].

Cusack & Roberts [11] studied sequential streaming using stimuli similar to those of Dannenbring & Bregman [43]. In one experiment, they used an interleaved melody task in which subjects had to detect a small change in the melody produced by a sequence of target tones in the presence of interleaved distracting tones. When the target and distracting tones fell in the same spectral region, performance was better when the target and distracting tones differed in their envelopes (tone versus noise) than when they had similar envelopes. This suggests that differences in envelope can enhance stream segregation under conditions where segregation is advantageous.

In a second experiment, Cusack & Roberts [11] used a rhythm discrimination task requiring judgements of the relative timing of successive sounds. Two sounds, X and Y, were alternated, with brief silent intervals between them. The difference in centre frequency of X and Y required to reach the threshold for detecting an irregularity in tempo of X and Y was measured; the threshold should be lower under conditions where envelope differences between X and Y contribute to obligatory segregation. They found that the threshold was about 30 per cent lower when there were differences in envelope of successive sounds (tone–noise) than when successive sounds had similar envelopes (tone–tone or noise–noise). This suggests that envelope differences contribute to obligatory stream segregation.

Grimault *et al.* [45] studied the perception of ABA–ABA sequences, where A and B were bursts of

broadband noise that were 100 per cent amplitude-modulated. The modulation rate of A was fixed at 100 Hz and the modulation rate of B was varied. Listeners generally perceived the sequences as a single stream when the difference in amplitude modulation rate was less than 0.75 oct, and as two streams when the difference was greater than about 1.0 oct. These results indicate that stream segregation can be produced by large differences in envelope rate.

Iverson [46] examined the influence of timbre on auditory streaming; timbre can be affected by spectral shape, changes in spectrum over time and envelope shape [47–49]. In his first experiment, listeners heard sequences of orchestral tones equated for pitch and loudness, and they rated how strongly the instruments segregated. Multi-dimensional scaling analyses of these ratings revealed that segregation was influenced by the envelopes of the tones, as well as by differences in spectral shape; tones with dissimilar envelopes received higher segregation ratings than tones with similar envelopes. In a second experiment, listeners heard interleaved melodies and tried to recognize the melodies played using a target timbre. The results showed better performance when successive sound differed in the rapidity of their onsets, although static spectral differences also played a role.

Singh & Bregman [50] investigated the role of harmonic content, envelope shape and F0 on streaming. Stimuli were presented in an ABA–ABA sequence. A and B were both complex tones that could differ in harmonic content (containing either the first two or first four harmonics) and in envelope shape (fast rise/slow fall or vice versa). The F0 difference between the tones was progressively increased and listeners had to indicate when they no longer heard the gallop rhythm. The F0 difference at which fusion ceased was least when both envelope and harmonic content were different between A and B and remained small when only harmonic content differed. The F0 difference at which fusion ceased was largest when A and B had the same envelope and harmonic content, and was somewhat less when only the envelopes differed, indicating a role for the envelope in streaming.

In summary, differences in envelope between successive sounds increase the tendency to hear segregation and lead to increased segregation under conditions both where segregation is advantageous and where it is disadvantageous.

(b) Segregation based on difference in phase spectrum

Roberts et al. [51] studied the effect of phase spectrum on streaming. The stimuli were harmonic complex tones with F0 = 100 Hz, filtered so as to contain only high harmonics, which are not resolved, i.e. which cannot be heard as separate tones [52]. To ensure that the results would not be influenced by combination tones, whose level can be affected by the relative phases of the primary components [53,54], a continuous lowpass noise was presented together with the complex tones. One tone, C, had harmonics added in cosine phase, which means that at the start of the sound, all of the harmonics started at an amplitude
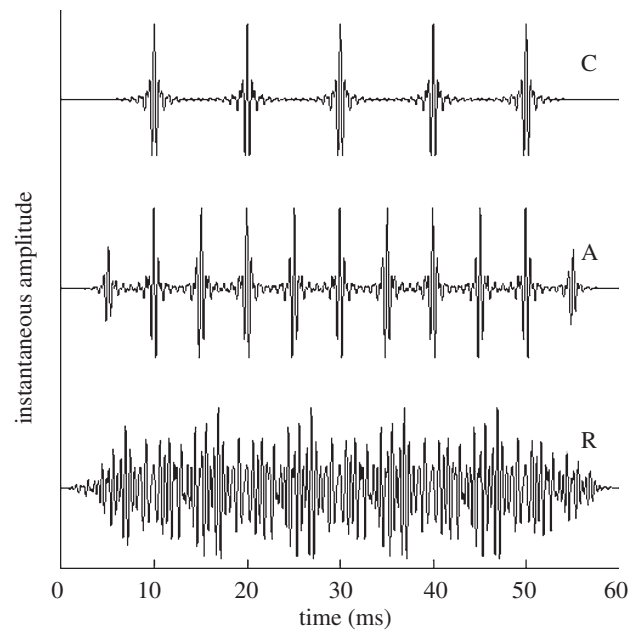

Figure 3. Waveforms of the stimuli used by Roberts et al. [51]. All stimuli were harmonic complex tones with F0 = 100 Hz. The components were added in cosine phase (C – top), alternating phase (A – middle) or random phase (R – bottom). Waveforms are shown before filtering. In the experiment, the stimuli were filtered so as to contain only high harmonics.

peak. This leads to a waveform with one large peak per period (see figure 3, top trace); the resulting sound has a sharp 'buzzy' timbre and a clear pitch. Another tone, A, had successive harmonics added in alternating cosine and sine phase, which means that at the start of the sound, the odd harmonics (1, 3, 5 . . .) started at an amplitude peak while the even harmonics (2, 4, 6 . . .) started at a positive-going zero-crossing. This leads to a waveform with two major peaks per period (figure 3, middle trace); the resulting sound has a clear pitch, which is roughly one octave higher than for tone C [55,56], and a less 'buzzy' timbre. Finally, a tone R had harmonics added in random phase, which leads to a more noise-like quality and a less clear pitch (figure 3, bottom trace). Roberts et al. filtered the stimuli into three different frequency regions, referred to as 'low', 'medium' and 'high', although in all cases the spectral regions were sufficiently high to prevent resolution of individual harmonics.

In one experiment, Roberts et al. [51] used subjective judgements of streaming to examine the perception of sequences of the form CAC–CAC or CRC–CRC. The tones C and A were either filtered into the same frequency region (high–high or low–low) or into different frequency regions. Subjects were asked to indicate continuously whether they heard a single stream or two streams as the 30 s sequence progressed. When the tones were filtered in the same frequency region, segregation was much more prevalent for the sequences CAC–CAC and CRC–CRC than for a repeating sequence of the same tone (CCC–CCC). The difference was especially large for the CRC–CRC sequence. Even when the tones were filtered into different frequency regions, stream segregation was significantly more prevalent for the CAC–CAC and CRC–CRC sequences than for the CCC–CCC sequence.

In a second experiment, Roberts *et al.* [51] used a rhythm-discrimination task requiring judgement of the relative timing of successive sounds. When the tones were filtered into the same frequency region, performance was much poorer when successive sounds differed in phase spectrum (CACAC... or CRCRC...) than when they had the same phase spectrum (CCCCC...). Thus, differences in phase spectrum can produce obligatory stream segregation for sounds with identical power spectra. The effects of phase are presumably mediated by changes in the waveform or envelope of the sound produced by the phase manipulation.

(c) Segregation based on fundamental frequency

Singh [57] and Bregman *et al.* [58] both used tone sequences where successive tones could differ either in spectral envelope, in F0, or both. They found that F0 and spectral shape can contribute independently to stream segregation. But, as both experiments used tones consisting of resolved harmonics, it is possible that the effect of F0 was mediated by changes in place of excitation of individual harmonics.

Vliegen & Oxenham [59] used sequences of complex tones consisting of high, unresolved harmonics with a variable F0 difference between successive tones, while keeping the spectral envelope the same. For these stimuli, the excitation patterns of successive tones would have been very similar. A task requiring subjective judgements of segregation showed that listeners segregated the tone sequences on the basis of F0 differences (presumably conveyed by temporal information alone), and that the pattern of judgements did not differ from that obtained with sequences of pure tones. However, using a similar task, Grimault *et al.* [60] found that stream segregation based on F0 differences of successive tones was stronger for complex tones containing resolved harmonics than for tones containing only unresolved harmonics. In a second experiment, Vliegen & Oxenham [59] showed that listeners could also use F0 differences to recognize a short atonal melody interleaved with random distracting tones if the melody was in an F0 range 11 semitones lower than that of the distracting tones. Thus, listeners can use large F0 differences between successive tones to achieve segregation.

Vliegen *et al.* [12] used a temporal discrimination task, in which best performance would be achieved if successive tones were not heard in separate perceptual streams. An ABA–ABA sequence was used, in which tone B could be either exactly at the temporal midpoint between two successive A tones or slightly delayed. The smallest detectable temporal shift was called the 'shift threshold'. The tones A and B were of three types: (i) both pure tones, tone A having a fixed frequency of 300 Hz; (ii) both complex tones filtered with a fixed passband so as to contain only high unresolvable harmonics, where only F0 was varied between tones A and B, and the F0 for tone A was 100 Hz; and (iii) both complex tones with the same F0 (100 Hz), but where the centre frequency of the passband varied between tones. For all three conditions, shift thresholds increased with increasing interval between tones A and B, but the effect was largest for the conditions where A and B differed in spectrum (i.e. the pure-tone and the variable-centre-frequency conditions). The results suggest that spectral information is dominant in inducing obligatory segregation, but periodicity (F0) information can also play a role.

The smaller effect found when A and B differed only in F0 may have been caused by the relatively small perceptual differences produced by the differences in F0. For complex tones filtered to contain only high harmonics, thresholds for discriminating changes in F0 are much higher than when the tones contain low harmonics [61–63].

In summary, differences in F0 alone can be as potent as spectral differences in promoting stream segregation, when segregation is advantageous, but F0 differences are less potent than spectral differences in producing obligatory segregation, when fusion is advantageous.

(d) Segregation based on lateralization

One extreme form of 'peripheral channelling' can be produced by presenting successive sounds to opposite ears. We describe this as a difference between the sounds in the 'ear of entry'. However, it is also possible to produce the percept of a sound alternating between the two ears by manipulating the ITD. For example, if sound A leads at the left ear by 500 μs, and sound B leads at the right ear by 500 μs, the sequence ABABAB is heard with the sound A towards the left ear and the sound B towards the right ear. We describe this as a difference in 'perceived location', produced without a difference in the ear of entry. If stream segregation can be produced by manipulation of ITDs, this should not be described as *peripheral* channelling, because lateralization depends on processing in the brainstem and higher levels of the auditory system.

Studies on the role of lateralization in stream segregation have shown that both ear of entry and perceived location (in the absence of differences in the ear of entry) affect sequential sound segregation, although ear of entry seems to be somewhat more important [64]. Hartmann & Johnson [6] found that performance in a melody recognition task with two interleaved melodies was nearly as good when the alternating tones of the two melodies were *perceived* in opposite ears owing to ITDs as when they were actually presented to opposite ears. In both conditions, performance was much better than when all tones were presented to one ear.

Gockel *et al.* [65] measured the threshold for detecting changes in F0 (called the F0DL) of a 100 ms monaural target complex tone in the presence and absence of preceding and following 'fringes', which were 200 ms harmonic complex tones. The F0 of the target differed across the intervals of a forced-choice trial, while the F0 of the fringes was fixed. It was argued that, if the fringes are perceptually grouped with the target, then they may interfere with F0 discrimination, while if they are segregated from the target, there should be little interference. The nominal F0 was 88 or 250 Hz, and was either the same or different for the target and fringes. The target and

fringes were filtered into a low- (125–625 Hz), a mid- (1375–1875 Hz) or a high-frequency (3900–5400 Hz) region, and this frequency region was either the same or different for the target and fringes. For the low-frequency region, resolved harmonics would have been present for both nominal F0s [66]. For the mid-frequency region, resolved harmonics would have been present for the 250 Hz F0 but not for the 88 Hz F0. For the high-frequency region, no resolved harmonics would have been present for either F0.

In some cases, presentation of the fringes to the same ear as the target markedly increased F0DLs relative to those observed in the absence of any fringes. For the impairment to occur, the target and fringes had to be in the same frequency region. Also, if all harmonics of the target and fringes were unresolved, then the impairment occurred even when the target and fringes differed in F0; otherwise, the impairment occurred only when the target and fringes had the same nominal F0. The impairment was substantially reduced when the perceived location of the fringes was shifted away from that of the target tones. This was true for shifts in perceived location produced by either interaural level differences (ILDs) or ITDs.

These findings are consistent with the idea that the fringes produced interference in F0 discrimination when the listeners had difficulty in segregating the target from the fringes. Segregation was promoted by: a difference between the target and fringes in spectral region; a difference in F0 when some harmonics were resolved; and a difference in perceived location.

Another study of segregation due to ITD was conducted by Sach & Bailey [67]. They presented tone sequences defining one of two simple target rhythms, interleaved with arrhythmic masking tones. The listeners' task was to attend to and identify the target rhythm. For one condition, the ILD of the target tones was 0 dB, while for the other the ILD was 4 dB. The ITD of the target tones was always zero. For the masking tones, the ILD was 0 dB and ITDs were varied. The results showed that target rhythm identification accuracy was low when the target and the masker had the same ILD and ITD, indicating that the maskers were effective when the target and the masker shared spatial position. Differences in ITD between the target and masking tones led to better accuracy, even when there were no differences in ILD, supporting the idea that ITDs can lead to stream segregation under conditions where segregation is advantageous.

Boehnke & Phillips [68] examined whether ITDs could induce obligatory stream segregation. In one experiment, they measured the ability to detect a temporal offset of the B sounds in a repeating ABA–ABA sequence. The sequence was preceded by three A sounds to promote the build-up of stream segregation, and the same temporal offset was present in all the ABA triplets. Conditions were included where both A and B were diotic noise bursts, and where A and B were noise bursts with opposite ITDs of ±500 µs. Performance did not differ significantly for the two conditions, suggesting that the ITDs did not induce obligatory segregation of the A and B noise bursts. However, it is not clear whether the preceding A tones would have been sufficient to produce a full build-up of stream segregation prior to presentation of the first ABA triplet.

Stainsby *et al.* [69] also investigated whether ITDs can lead to obligatory segregation, using a rhythm-discrimination task. A relatively long inducer sequence was used before introducing the irregularity in rhythm that was to be detected. This allowed the tendency for segregation to build up. Stimuli were bandpass-filtered harmonic complex tones with F0 = 100 Hz. The alternating A and B tones had equal but opposite ITDs of 0, 0.25, 0.5, 1 or 2 ms and had the same or different passbands. The passband ranges were 1250–2500 Hz and 1768–3536 Hz in experiment 1, and 353–707 Hz and 500–1000 Hz in experiment 2. In both experiments, increases in ITD led to increases in threshold, mainly when the passbands of A and B were the same. However, the effects were largest for ITDs above 0.5 ms, for which rhythmic irregularities in the timing of the tones within each ear may have disrupted performance. The effects were small for ITDs of up to 0.5 ms. Stainsby *et al.* concluded that differences in apparent spatial location produced by ITD have only weak effects on obligatory streaming.

Darwin & Hukin [70] showed that sequential grouping for speech can be strongly influenced by ITD. In one experiment, they simultaneously presented two sentences as illustrated in figure 4. They varied the ITDs of the two sentences, in the range 0 to ±181 µs. For example, one sentence might lead in the left ear by 45 µs, while the other sentence would lead in the right ear by 45 µs. The sentences were based on natural speech but were processed so that each was spoken on a monotone, i.e. with constant F0. The F0 difference between the two sentences was varied from 0 to 4 semitones. Subjects were instructed to attend to one particular sentence. At a certain point, the two sentences contained two different semantically plausible target words, aligned in starting time and duration. The F0s and the ITDs of the two target words were varied independently from those of the two sentences. Subjects had to indicate which of the two target words they heard in the attended sentence. They reported the target word that had the same ITD as the attended sentence much more often than the target word with the opposite ITD. In other words, the target word with the same ITD as the attended sentence was grouped with that sentence. This was true even when the target word had the same ITD as the attended sentence but a different F0. Thus, subjects grouped words across time according to their perceived location, independent of F0 differences (for differences up to four semitones).

In a second experiment, Darwin & Hukin [70] showed that listeners did not explicitly track frequency components that shared a common ITD. They manipulated the ITD of a single harmonic close in frequency to the first formant in a vowel, and showed that the manipulation had little effect on the vowel identity, either when the vowel was presented alone, or when it was presented in a 'carrier' sentence with the same ITD as the main part of the vowel. Darwin & Hukin [70] concluded that listeners who try to track a
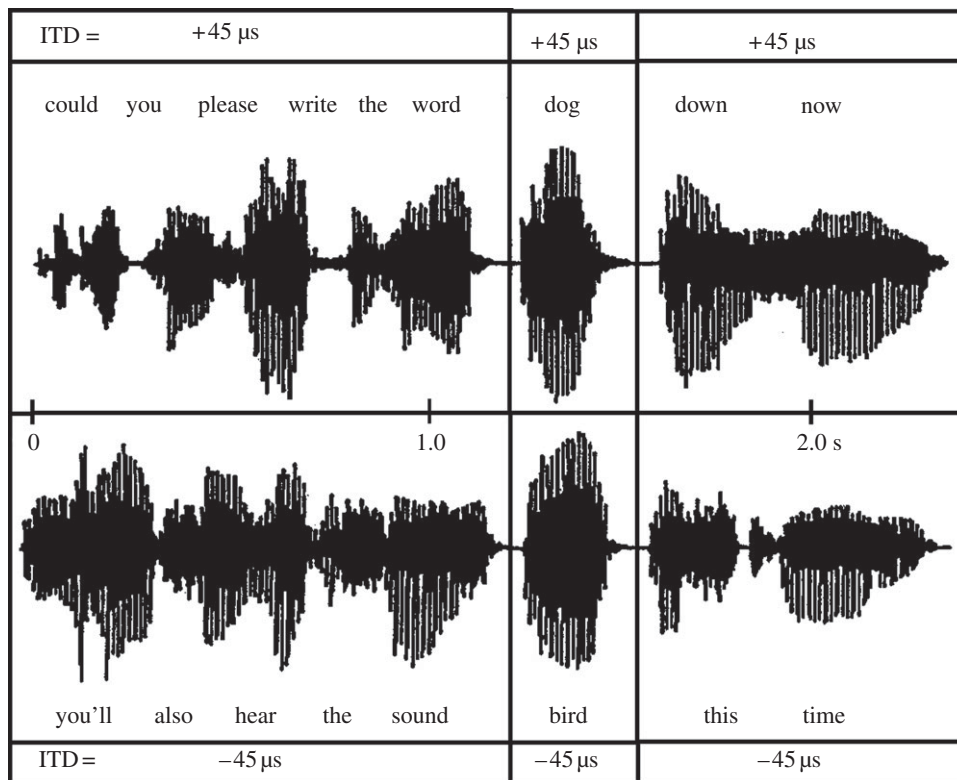
Figure 4. Example of the stimuli used by Darwin & Hukin [70]. See text for details.

particular sound source over time direct attention to auditory objects at a particular subjective location, and not to frequency components that share a common ITD. The auditory objects themselves may be formed using cues other than ITD, for example onset and offset asynchrony [71], harmonicity [72] and ear of entry [71].

(e) The role of pattern regularity in stabilizing streams

In most of the experiments described so far involving sequences of A and B tones, the A tones were all the same and the B tones were all the same. However, it is possible to introduce patterns into the A tones alone, for example, by adding an accent (an increase in level) to some tones or changing the frequency of some tones slightly. These patterns may be regular (e.g. an accent placed on every third A tone) or irregular. The possible influence of such patterns on streaming was explored by Bendixen *et al.* [4]. Listeners were asked to indicate continuously whether they perceived a sequence of tones as one stream or two. In some conditions, regular patterns were present in either the A tones or the B tones, or both. As typically occurs, the percept tended to flip between one stream and two streams, i.e. bistability occurred. However, regular patterns in either the A or the B tones, or both, prolonged the mean duration of two-stream percepts relative to the case where the patterns were irregular, whereas the duration of one-stream percepts was unaffected. The authors suggested that temporal regularities stabilize streams once they have been formed on the basis of simpler acoustic cues. These regularities occur over fairly long durations and their

detection and utilization presumably depend on relatively central processes.

Regularity in the timing of the tones can also influence stream formation. Devergie *et al.* [14] studied the ability to identify familiar melodies in an interleaved-melody task and found that identification was better when the rhythm of the distractor tones was regular than when it was irregular.

7. CONCLUSIONS

When the differences between successive sounds in a rapid sequence are small, the sequence tends to be heard as a single stream. When the differences are large, two or more streams are usually perceived. For intermediate differences, the percept tends to flip often, but not in a regular manner, between fusion and segregation, a property called bistability.

Stream segregation shows a tendency to build up over time. The build-up can be partially reset by a sudden change in the sequence and/or by switches in attention. There is controversy as to whether or not the build-up of segregation requires attention to the sound sequence.

Differences in the excitation patterns evoked by the different sounds in a sequence have a strong influence on stream segregation. This is often described as 'peripheral channelling'. However, factors other than peripheral channelling can play a strong role; stream segregation can occur when successive sounds are presented to the same ear and have essentially identical excitation patterns. Furthermore, stream segregation produced by these other factors can be obligatory, i.e. it occurs even in tasks where better performance would be achieved if successive sounds could be

perceived as fused. The factors other than differences in excitation patterns that can influence streaming include: differences in envelope; differences in F0; differences in phase spectrum (which can affect the waveform and hence the pitch and subjective quality of sounds); differences in lateralization produced by ITDs or ILDs; and regularities of the patterns within perceived streams.

In some cases, the sequential stream segregation produced by these other factors can be as strong as that produced by differences in power spectrum (and hence excitation patterns). However, in other cases, the 'other factors' produce weaker effects. The potency of a given factor in producing stream segregation may be related to the perceptual salience of changes produced by manipulating that factor and may depend on the task used to measure segregation. For example, differences in ITD seem to be effective in producing segregation when it is advantageous (e.g. in a task involving the recognition of interleaved melodies), but differences in ITD are not very effective in producing obligatory segregation, even though their effects are subjectively salient.

REFERENCES

1 Bregman, A. S. 1990 *Auditory scene analysis: the perceptual organization of sound*. Cambridge, MA: Bradford Books, MIT Press.

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

3 Pressnitzer, D. & Hupé, J. M. 2006 Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Curr. Biol.* **16**, 1351–1357. (doi:10.1016/j.cub.2006.05.054)

4 Bendixen, A., Denham, S. L., Gyimesi, K. & Winkler, I. 2010 Regular patterns stabilize auditory streams. *J. Acoust. Soc. Am.* **128**, 3658–3666. (doi:10.1121/1.3500695)

5 Dowling, W. J. 1973 The perception of interleaved melodies. *Cognitive Psychol.* **5**, 322–337. (doi:10.1016/0010-0285(73)90040-6)

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

7 Broadbent, D. E. & Ladefoged, P. 1959 Auditory perception of temporal order. *J. Acoust. Soc. Am.* **31**, 151–159.

8 Warren, R. M., Obusek, C. J., Farmer, R. M. & Warren, R. P. 1969 Auditory sequence: confusion of patterns other than speech or music. *Science* **164**, 586–587. (doi:10.1126/science.164.3879.586)

9 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. (doi:10.1037/h0031163)

10 Cusack, R. & Roberts, B. 1999 Effects of similarity in bandwidth on the auditory sequential streaming of two-tone complexes. *Perception* **28**, 1281–1289. (doi:10.1068/p2804)

11 Cusack, R. & Roberts, B. 2000 Effects of differences in timbre on sequential grouping. *Percept. Psychophys.* **62**, 1112–1120. (doi:10.3758/BF03212092)

12 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. (doi:10.1121/1.427140)

13 Bey, C. & McAdams, S. 2002 Schema-based processing in auditory scene analysis. *Percept. Psychophys.* **64**, 844–854. (doi:10.3758/BF03194750)

14 Devergie, A., Grimault, N., Tillmann, B. & Berthommier, F. 2010 Effect of rhythmic attention on the segregation of interleaved melodies. *J. Acoust. Soc. Am.* **128**, EL1–EL7. (doi:10.1121/1.3436498)

15 Bregman, A. S. 1978 Auditory streaming is cumulative. *J. Exp. Psychol.: Human Percept. Perf.* **4**, 380–387. (doi:10.1037/0096-1523.4.3.380)

16 Anstis, S. & Saida, S. 1985 Adaptation to auditory streaming of frequency-modulated tones. *J. Exp. Psychol.: Human Percept. Perf.* **11**, 257–271. (doi:10.1037/0096-1523.11.3.257)

17 Denham, S. L. & Winkler, I. 2006 The role of predictive models in the formation of auditory streams. *J. Physiol. Paris* **100**, 154–170. (doi:10.1016/j.jphysparis.2006.09.012)

18 Bregman, A. S., Ahad, P. A., Crum, P. A. & O'Reilly, J. 2000 Effects of time intervals and tone durations on auditory stream segregation. *Percept. Psychophys.* **62**, 626–636. (doi:10.3758/BF03212114)

19 Rogers, W. L. & Bregman, A. S. 1993 An experimental evaluation of three theories of auditory stream segregation. *Percept. Psychophys.* **53**, 179–189. (doi:10.3758/BF03211728)

20 Rogers, W. L. & Bregman, A. S. 1998 Cumulation of the tendency to segregate auditory streams: resetting by changes in location and loudness. *Percept. Psychophys.* **60**, 1216–1227. (doi:10.3758/BF03206171)

21 Roberts, B., Glasberg, B. R. & Moore, B. C. J. 2008 Effects of the build-up and resetting of auditory stream segregation on temporal discrimination. *J. Exp. Psychol.: Human Percept. Perf.* **34**, 992–1006. (doi:10.1037/0096-1523.34.4.992)

22 Haywood, N. R. & Roberts, B. 2010 Build-up of the tendency to segregate auditory streams: resetting effects evoked by a single deviant tone. *J. Acoust. Soc. Am.* **128**, 3019–3031. (doi:10.1121/1.3488675)

23 Beauvois, M. W. & Meddis, R. 1997 Time decay of auditory stream biasing. *Percept. Psychophys.* **59**, 81–86. (doi:10.3758/BF03206850)

24 Cusack, R., Deeks, J., Aikman, G. & Carlyon, R. P. 2004 Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *J. Exp. Psychol.: Human Percept. Perf.* **30**, 643–656. (doi:10.1037/0096-1523.30.4.643)

25 Snyder, J. S., Carter, O. L., Lee, S. K., Hannon, E. E. & Alain, C. 2008 Effects of context on auditory stream segregation. *J. Exp. Psychol.: Human Percept. Perf.* **34**, 1007–1016. (doi:10.1037/0096-1523.34.4.1007)

26 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.: Human Percept. Perf.* **27**, 115–127. (doi:10.1037/0096-1523.27.1.115)

27 Carlyon, R. P., Plack, C. J., Fantini, D. A. & Cusack, R. 2003 Cross-modal and non-sensory influences on auditory streaming. *Perception* **32**, 1393–1402. (doi:10.1068/p5035)

28 Moore, B. C. J. & Gockel, H. 2002 Factors influencing sequential stream segregation. *Acta Acust.-Acust.* **88**, 320–333.

29 Thompson, S. K., Carlyon, R. P. & Cusack, R. 2011 An objective measurement of the build-up of auditory

streaming and of its modulation by attention. *J. Exp. Psychol.: Human Percept. Perf.* **37**, 1253–1262. (doi:10.1037/a0021925)

30 Salame, P. & Baddeley, A. D. 1982 Disruption of short-term memory by unattended speech: implications for the structure of working memory. *J. Verb. Learn. Verb. Behav.* **21**, 150–164. (doi:10.1016/S0022-5371(82)90521-7)

31 Jones, D., Alford, D., Bridges, A., Tremblay, S. & Macken, B. 1999 Organizational factors in selective attention: the interplay of acoustic distinctiveness and auditory streaming in the irrelevant sound effect. *J. Exp. Psychol.: Learn., Mem., Cog.* **25**, 464–473. (doi:10.1037/0278-7393.25.2.464)

32 Macken, W. J., Tremblay, S., Houghton, R. J., Nicholls, A. P. & Jones, D. M. 2003 Does auditory streaming require attention? Evidence from attentional selectivity in short-term memory. *J. Exp. Psychol.: Human Percept. Perf.* **29**, 43–51. (doi:10.1037/0096-1523.29.1.43)

33 Moore, B. C. J. & Glasberg, B. R. 1983 Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *J. Acoust. Soc. Am.* **74**, 750–753. (doi:10.1121/1.389861)

34 Moore, B. C. J. 2003 *An introduction to the psychology of hearing*, 5th ed. Bingley, UK: Emerald.

35 Beauvois, M. W. & Meddis, R. 1996 Computer simulation of auditory stream segregation in alternating-tone sequences. *J. Acoust. Soc. Am.* **99**, 2270–2280. (doi:10.1121/1.415414)

36 McCabe, S. L. & Denham, M. J. 1997 A model of auditory streaming. *J. Acoust. Soc. Am.* **101**, 1611–1621. (doi:10.1121/1.418176)

37 Haywood, N. R. & Roberts, B. 2011 Sequential grouping of pure-tone percepts evoked by the segregation of components from a complex tone. *J. Exp. Psychol.: Human Percept. Perf.* **37**, 1263–1274. (doi:10.1037/a0023416)

38 Hartmann, W. M. 1997 *Signals, sound, and sensation.* Woodbury, New York: AIP Press.

39 Rose, M. M. & Moore, B. C. J. 2000 Effects of frequency and level on auditory stream segregation. *J. Acoust. Soc. Am.* **108**, 1209–1214. (doi:10.1121/1.1287708)

40 Zwicker, E. 1961 Subdivision of the audible frequency range into critical bands (Frequenzgruppen). *J. Acoust. Soc. Am.* **33**, 248. (doi:10.1121/1.1908630)

41 Bregman, A. S., Ahad, P. A. & Van Loon, C. 2001 Stream segregation of narrow-band noise bursts. *Percept. Psychophys.* **63**, 790–797. (doi:10.3758/BF03194438)

42 Cusack, R. & Roberts, B. 2004 Effects of differences in the pattern of amplitude envelopes across harmonics on auditory stream segregation. *Hear. Res.* **193**, 95–104. (doi:10.1016/j.heares.2004.03.009)

43 Dannenbring, G. L. & Bregman, A. S. 1976 Stream segregation and the illusion of overlap. *J. Exp. Psychol.: Human Percept. Perf.* **2**, 544–555. (doi:10.1037/0096-1523.2.4.544)

44 Moore, B. C. J. & Glasberg, B. R. 1986 The role of frequency selectivity in the perception of loudness, pitch and time. In *Frequency selectivity in hearing* (ed. B. C. J. Moore), pp. 251–308. London, UK: Academic.

45 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. (doi:10.1121/1.1452740)

46 Iverson, P. 1995 Auditory stream segregation by musical timbre: effects of static and dynamic acoustic attributes. *J. Exp. Psychol.: Human Percept. Perf.* **21**, 751–763. (doi:10.1037/0096-1523.21.4.751)

47 Grey, J. M. 1977 Multidimensional perceptual scaling of musical timbres. *J. Acoust. Soc. Am.* **61**, 1270–1277. (doi:10.1121/1.381428)

48 Krumhansl, C. L. 1989 Why is musical timbre so hard to understand? In *Structure and perception of electroacoustic sound and music* (eds S. Nielzén & O. Olssen), pp. 43–53. Amsterdam, The Netherlands: Excerpta Medica.

49 Iverson, P. & Krumhansl, C. L. 1993 Isolating the dynamic attributes of musical timbre. *J. Acoust. Soc. Am.* **94**, 2595–2603. (doi:10.1121/1.407371)

50 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. (doi:10.1121/1.419688)

51 Roberts, B., Glasberg, B. R. & Moore, B. C. J. 2002 Primitive stream segregation of tone sequences without differences in F0 or passband. *J. Acoust. Soc. Am.* **112**, 2074–2085. (doi:10.1121/1.1508784)

52 Moore, B. C. J. & Gockel, H. 2011 Resolvability of components in complex tones and implications for theories of pitch perception. *Hear. Res.* **276**, 88–97. (doi:10.1016/j.heares.2011.01.003)

53 Buunen, T. J. F., Festen, J. M., Bilsen, F. A. & van den Brink, G. 1974 Phase effects in a three-component signal. *J. Acoust. Soc. Am.* **55**, 297–303. (doi:10.1121/1.1914501)

54 Pressnitzer, D. & Patterson, R. D. 2001 Distortion products and the pitch of harmonic complex tones. In *Physiological and psychophysical bases of auditory function* (eds D. J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs & R. Schoonhoven), pp. 97–103. Maastricht, The Netherlands: Shaker.

55 Patterson, R. D. 1987 A pulse ribbon model of monaural phase perception. *J. Acoust. Soc. Am.* **82**, 1560–1586. (doi:10.1121/1.395146)

56 Shackleton, T. M. & Carlyon, R. P. 1994 The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *J. Acoust. Soc. Am.* **95**, 3529–3540. (doi:10.1121/1.409970)

57 Singh, P. G. 1987 Perceptual organization of complex-tone sequences: a tradeoff between pitch and timbre? *J. Acoust. Soc. Am.* **82**, 886–895. (doi:10.1121/1.395287)

58 Bregman, A. S., Liao, C. & Levitan, R. 1990 Auditory grouping based on fundamental frequency and formant peak frequency. *Can. J. Psychol.* **44**, 400–413. (doi:10.1037/h0084255)

59 Vliegen, J. & Oxenham, A. J. 1999 Sequential stream segregation in the absence of spectral cues. *J. Acoust. Soc. Am.* **105**, 339–346. (doi:10.1121/1.424503)

60 Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P. & Collet, L. 2000 Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency. *J. Acoust. Soc. Am.* **108**, 263–271. (doi:10.1121/1.429462)

61 Hoekstra, A. & Ritsma, R. J. 1977 Perceptive hearing loss and frequency selectivity. In *Psychophysics and physiology of hearing* (eds E. F. Evans & J. P. Wilson), pp. 263–271. London, UK: Academic Press.

62 Moore, B. C. J. & Glasberg, B. R. 1988 Effects of the relative phase of the components on the pitch discrimination of complex tones by subjects with unilateral cochlear impairments. In *Basic issues in hearing* (eds H. Duifhuis, H. Wit & J. Horst), pp. 421–430. London, UK: Academic Press.

63 Houtsma, A. J. M. & Smurzynski, J. 1990 Pitch identification and discrimination for complex tones with many harmonics. *J. Acoust. Soc. Am.* **87**, 304–310. (doi:10.1121/1.399297)

64 Gockel, H. 2000 Perceptual grouping and pitch perception. In *Results of the 8th Oldenburg Symposium on Psychological Acoustics* (eds A. Schick, M. Meis & C. Reckhardt), pp. 275–292. Oldenburg, Germany: BIS.

65 Gockel, H., Carlyon, R. P. & Micheyl, C. 1999 Context dependence of fundamental-frequency discrimination: lateralized temporal fringes. *J. Acoust. Soc. Am.* **106**, 3553–3563. (doi:10.1121/1.428208)

66 Plomp, R. 1964 The ear as a frequency analyzer. *J. Acoust. Soc. Am.* **36**, 1628–1636. (doi:10.1121/1.1919256)

67 Sach, A. J. & Bailey, P. J. 2004 Some characteristics of auditory spatial attention revealed using rhythmic masking release. *Percept. Psychophys.* **66**, 1379–1387. (doi:10.3758/BF03195005)

68 Boehnke, S. E. & Phillips, D. P. 2005 The relation between auditory temporal interval processing and sequential stream segregation examined with stimulus laterality differences. *Percept. Psychophys.* **67**, 1088–1101. (doi:10.3758/BF03193634)

69 Stainsby, T. H., Füllgrabe, C., Flanagan, H. J., Waldman, S. & Moore, B. C. J. 2011 Sequential streaming due to manipulation of interaural time differences. *J. Acoust. Soc. Am.* **130**, 904–914. (doi:10.1121/1.3605540)

70 Darwin, C. J. & Hukin, R. W. 1999 Auditory objects of attention: the role of interaural time differences. *J. Exp. Psychol.: Human Percept. Perf.* **25**, 617–629. (doi:10.1037/0096-1523.25.3.617)

71 Darwin, C. J. & Carlyon, R. P. 1995 Auditory grouping. In *Hearing* (ed. B. C. J. Moore), pp. 387–424. San Diego, CA: Academic Press.

72 Moore, B. C. J., Glasberg, B. R. & Peters, R. W. 1986 Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *J. Acoust. Soc. Am.* **80**, 479–483. (doi:10.1121/1.394043)