

Auditory Stream Segregation (Sequential Integration)

David Meredith

Department of Computing,

City University, London.

`dave@titanmusic.com`

`www.titanmusic.com`

MSc/Postgraduate Diploma in Music Information Technology Lecture

Department of Music, City University, London

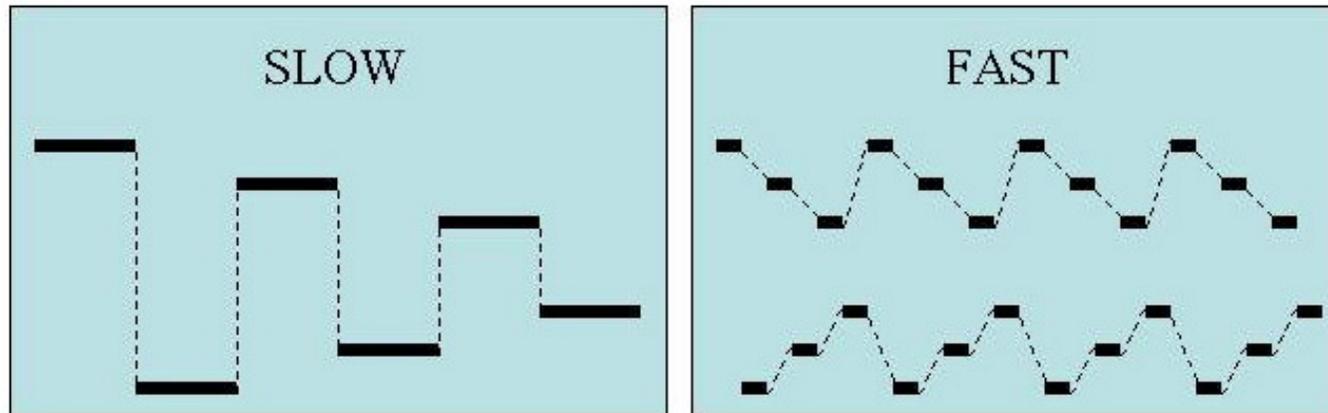
Friday, 14 March 2003.

1. Sequential integration

- *Sequential integration* is the connection of parts of an auditory spectrum over time to form concurrent streams (Bregman and Ahad, 1995, p. 7).
- The connection of consecutive tones played on a single instrument to form a single percept (stream) that we call a melody is an example of sequential integration.
- Another example of sequential integration occurs when we hear a sound to continue even when it is joined by other sounds to form a mixture.
- Sequential integration occurs until a sound changes suddenly, for example, in frequency content, timbre, fundamental, amplitude or spatial location.
- Compare with the way the Gestalt principle of similarity governs the way we segment musical passages into groups: sudden changes induce us to hear group boundaries (e.g., Lerdahl and Jackendoff's (1983, p. 46) GPR 3).
- Grouping is the segmentation of streams into structural units.
- In everyday sounds, as opposed to musical ones, we associate a separate stream with each separate sound source.
- When the sound of our environment reaches our ear, it arrives as a mixture of sounds from many sources.
- Our brain must process this mixed sound, identify the various sources of the sounds forming the mixture and then assign a stream to each source.
- Each stream can have its own melody and rhythm.
- We are much better at recognizing patterns of sounds and temporal relationships between sounds when the sounds are all in the same stream.

1. Sequential integration

2. Stream segregation in a cycle of six tones (Bregman and Ahad, 1995, p. 8, Track 1)

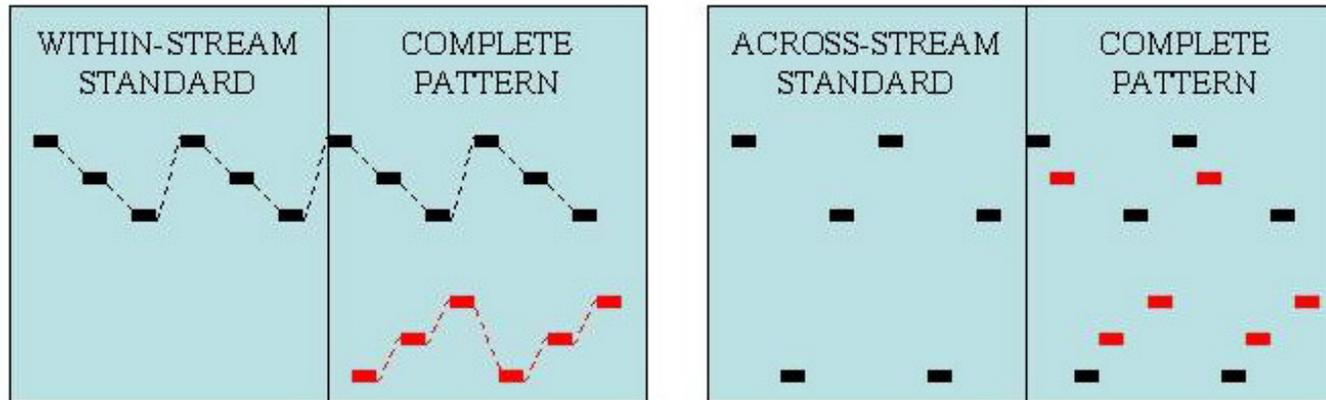


- Based on experiment by Bregman and Campbell (1971).
- When played slowly, all six tones integrate into a single stream.
- When played fast, pattern divides into two streams with the three high tones in one stream and the three low tones in the other.
- In slow version, easy to hear that first low tone comes immediately after first high tone.
- But when played fast, difficult to perceive temporal relationships between tones in different streams.

2. Stream segregation in a cycle of six tones
(Bregman and Ahad, 1995, p. 8, Track 1)

1. On their CD, Bregman and Ahad (1995) present 17 demonstrations that illustrate various aspects of auditory sequential integration.
2. The first of these is based on an experiment carried out by Bregman and Campbell (1971).
3. You'll hear a pattern of six tones, three high ones and three low ones. This pattern is repeated several times.
4. When the pattern is played slowly, you will hear all six tones as forming a single stream (voice, part, melody). This is indicated in the diagram on the left by the dotted lines that join each tone to the next.
5. However, when the pattern is played fast, you will hear two streams, one containing the high tones and the other containing the low tones.
6. In other words, when the melody is played fast, it sounds as though it splits into two parts.
7. [PLAY TRACK 1]
8. Note also how difficult it is to hear that the low tones in the fast version actually occur between the high tones—it is very difficult to hear precisely the temporal relationship between tones in different streams.

3. Pattern recognition, within and across perceptual streams (Bregman and Ahad, 1995, pp. 9–10, Track 2)

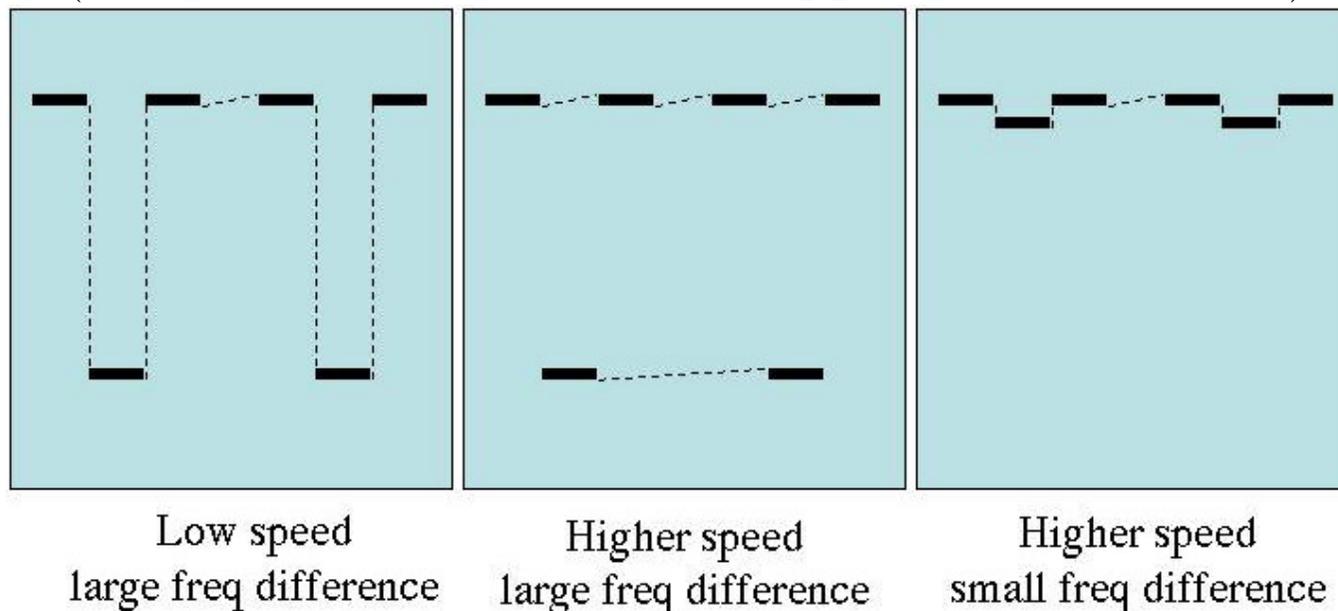


- Easy to hear ‘within-stream’ standard.
- Very hard to hear ‘across-stream’ standard.
- Shows that we find it hard to attend to more than one stream at a time.
- We also find it hard to switch our attention quickly back and forth between streams.

3. Pattern recognition, within and across perceptual streams
(Bregman and Ahad, 1995, pp. 9–10, Track 2)

1. In the previous demonstration, we saw that when the six-tone pattern was played fast, it was perceived to split into two streams.
2. In this demonstration, Bregman and Ahad (1995) show that when this happens, it is almost impossible for us to pay attention to both of the streams at once.
3. This example also shows that it is very difficult to switch our attention quickly from one stream to the other and then back again.
4. The demonstration is in two parts. In each part, you first hear a repeated three-tone standard pattern taken from the fast version of the six-tone pattern that you heard in the previous demonstration.
5. Then you hear the full six-tone pattern and you have to try to continue hearing the three-tone standard pattern within the full six-tone pattern.
6. In the first part of the demonstration, the three-tone pattern consists of the three high tones that are all heard to be integrated into a single stream in the full six-tone pattern (see left-hand figure).
7. In the second part of the demonstration, the three-tone standard contains tones that are in different streams in the complete six-tone pattern (see right-hand figure).
8. You should find that it is quite easy to hear the ‘within-stream’ standard pattern in the complete pattern but that it is very hard to hear the ‘across-stream’ standard in the complete six-tone pattern.
9. [PLAY TRACK 2]

4. The effects of speed and frequency on stream segregation (Bregman and Ahad, 1995, pp. 11–12, Track 3)



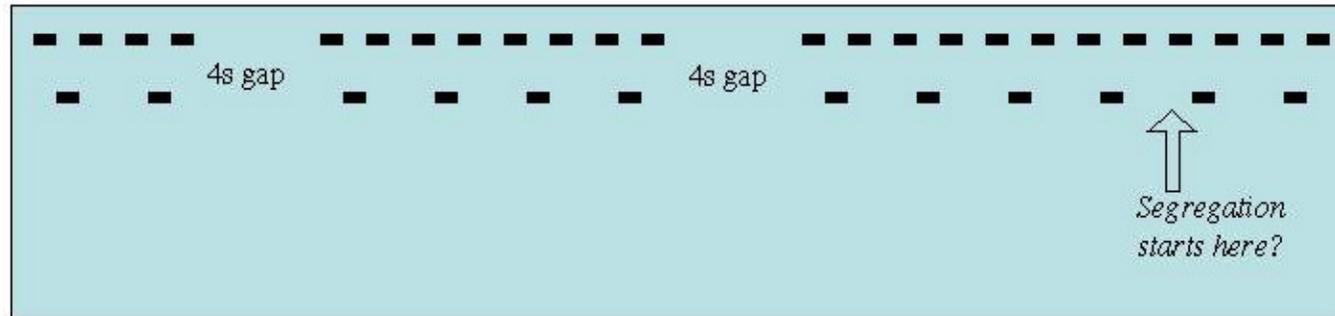
- When ‘galoping pattern’ (van Noorden, 1975, 1977) splits into two streams, each of these streams has a different isochronous rhythm.
- Each stream has its own rhythm and melody.
- In galoping pattern, a larger frequency difference between the high and low tones and a faster speed of presentation promote stream segregation.
- Result of Gestalt principle of proximity operating in a two-dimensional space in which one dimension is log-frequency and the other is time.

4. The effects of speed and frequency on stream segregation
(Bregman and Ahad, 1995, pp. 11–12, Track 3)

1. In his experiments on streaming, van Noorden (1975, 1977) used a repeated three-tone pattern consisting of two high tones with a lower tone in between as shown in the diagram on the right.
2. If the middle, lower tone in each occurrence of the pattern is sufficiently close in pitch and timbre to the two higher tones then the whole sequence is integrated into one stream with a so-called ‘galoping’ rhythm.
3. However, if the pitch of the middle tone in each occurrence of the pattern is lowered so that it is far away in pitch from the two higher tones, as shown in the middle diagram, then the sequence breaks into two streams, one containing the high tones and the other containing the low tones.
4. When this happens, the listener ceases to hear the galoping rhythm. Instead, each of the two streams is heard to have its own rhythm and the listener can attend to one or the other but not both simultaneously.
5. Each of these two streams has what is called an *isochronous* rhythm—that is, within each stream, the tones are equally spaced in time. However, the events in the upper stream occur twice as frequently as events in the lower stream.
6. So when the sequence is heard as one stream, it is heard to have a galoping rhythm which is quite different from the isochronous rhythms of the two streams that form when the sequence is segregated. This makes it particularly easy to identify the point at which the listener perceives the sequence to break up into two separate streams.
7. The melodic pattern of the sequence when it is integrated into one stream is also quite different from the melodic pattern of each of the two streams when the sequence is segregated. In the integrated sequence, we hear a repeated neighbour tone pattern, whereas each of the two streams that form when the sequence is segregated consists of simply a sequence of repeated tones.
8. This shows that each stream in an auditory field is perceived to have its own rhythm and melody.
9. If the frequency difference between the high and low tones in such a galoping pattern is held fixed and the speed at which the pattern is presented is gradually increased, then, to begin with, when the pattern is presented at a slow speed, the sequence is integrated into a single stream and a galoping pattern is heard, as shown here on the left.
10. However, when the speed of the sequence is increased, a point is reached at which the sequence is segregated into two streams, one containing the high tones and the other containing the low tones, and each of these streams will be perceived to have its own isochronous rhythm (see middle diagram).
11. In general, the greater the frequency difference between the high and low tones, the lower the speed required to split the sequence into two streams.

12. The demonstration I'm going to play you in a moment contains two examples.
13. In the first example, the frequency difference between the tones is large, as illustrated here in the diagrams on the left and in the middle, and the sequence is heard to split into two streams at a moderate speed.
14. However, in the second example, the tones are separated by just one semitone, as shown here in the diagram on the right, and you'll probably find that the sequence fails to segregate into two streams even at the highest speed.
15. [PLAY TRACK 3].
16. What seems to be happening here is that the Gestalt principle of proximity seems to be operating in a two-dimensional space in which one dimension is log frequency and the other is time.
17. The specific speeds at which sequences are heard to segregate into separate streams for a given frequency difference, tell us how to scale these axes so that the actual distances between events in the graphic representation can be used to predict, using the Gestalt principle of proximity, how the stimulus will be segregated into streams.

5. Effect of repetition on streaming (Bregman and Ahad, 1995, p. 13, Track 4)



- When we segregate a sequence into two streams, this does not happen immediately—it only happens after we've heard a few cycles of the repeating pattern.
- If we responded too quickly, this would give rise to highly volatile interpretations. We therefore accumulate evidence for a particular interpretation until we have enough to switch to that interpretation (Bregman, 1990, p. 130).
- As the sequence of repeated gallop patterns gets longer, our perception of segregation gets stronger.

5. Effect of repetition on streaming
(Bregman and Ahad, 1995, p. 13, Track 4)

1. You may have already noticed that you only perceive a repeating pattern sequence to segregate into separate streams after you have heard a few cycles of the sequence.
2. Bregman (1990, pp. 130) points out that if we split a stimulus into streams too quickly, we would be too sensitive to changes in evidence and we would have wildly oscillating interpretations of auditory scenes.
3. He therefore proposes that our auditory scene analysis system has evolved to be “sluggish” or ‘damped’ in its response to stimuli.
4. So when we hear a stimulus containing tones in two different frequency ranges, we start off by assuming the simplest explanation—that the two tones are coming from the same source. However, as we hear more and more tones clustered into two separate regions, the auditory system interprets this as greater and greater evidence that the tones are coming from two different sources and eventually switches to a two-stream interpretation.
5. In this demonstration you’re going to hear five sequences, each one containing a number of repetitions of a three-tone galloping pattern and each one twice as long as the previous one. The sequences are separated by 4s gaps.
6. In the first two sequences you will probably find that the sequence finishes before you segregate the galloping pattern into two isochronous streams.
7. As the sequences get longer, you will probably find that the perception of segregation gets stronger.
8. [PLAY TRACK 4]

6. Segregation of a melody from distractor tones (Bregman and Ahad, 1995, p. 14–15, Track 5)

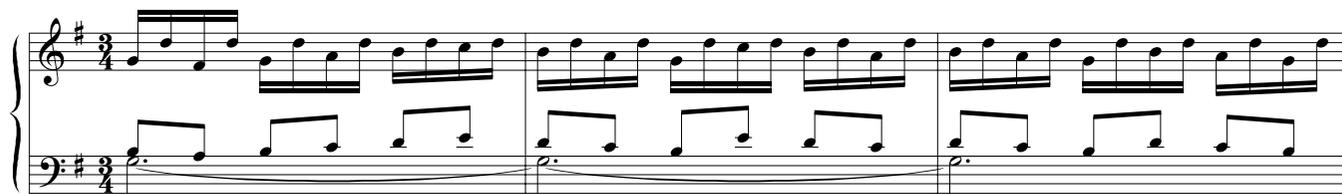
D I S M T E R L A O C D T Y O R S D I S^M T^E R^L A^O C^D T^Y O R S

- Segregation also occurs in non-repetitive sequences.
- Demonstration consists of 5 sequences.
- In each sequence, melody tones interleaved with distractor tones.
- In each successive sequence, frequency range of melody pulls further and further away from that of the distractors.
- In first sequence, melody frequency range is the same as that of the distractors therefore one stream is formed and the melody is hard to identify.
- In the fifth sequence, the melody frequency range is far from that of the distractors therefore two streams form and the melody is easy to identify.

6. Segregation of a melody from distractor tones
(Bregman and Ahad, 1995, p. 14–15, Track 5)

1. Stream segregation can also occur in sequences that do not consist of many repetitions of a short pattern.
2. In the demonstration I'm going to play to you in a moment, the sequence you'll hear is constructed by interleaving the tones of a familiar melody with random distractor tones in the way shown here on the left.
3. Each distractor tone is randomly chosen to be within 4 semitones of the previous melody tone.
4. You'll hear 5 versions of this tone sequence. The first time you hear it, both the melody tones and the distractors are in the same pitch range.
5. Each time the sequence is repeated, the melody tones are raised by two semitones but the distractor tones are left in the range of the original statement of the melody.
6. The frequency range of the melody therefore gradually pulls further and further away from that of the distractor tones.
7. In the first sequence in which the melody and distractor tones share the same frequency range, it is very hard to identify the melody. However, as the frequency range of the melody becomes more and more different from that of the distractors, it becomes easier and easier to identify.
8. In the first sequence where the melody and distractors share the same frequency range, all the tones in the sequence group into a single perceptual stream. Clearly, our auditory scene analysis system finds it difficult to form perceptual links between non-adjacent events within a stream, therefore we find it difficult to isolate the melody tones from the distractor tones.
9. However, as the frequency range of the melody becomes more and more different from that of the distractor tones, we find it easier and easier to segregate the sequence into two streams on the basis of frequency proximity, one stream containing the melody tones and the other containing the distractor tones.
10. Having segregated the melody from the distractor tones, we can form perceptual links between consecutive events in the melody stream and identify the melody.
11. [PLAY TRACK 5]

7. Segregation of high tones from low tones in a Prelude by Bach (Compound melody)

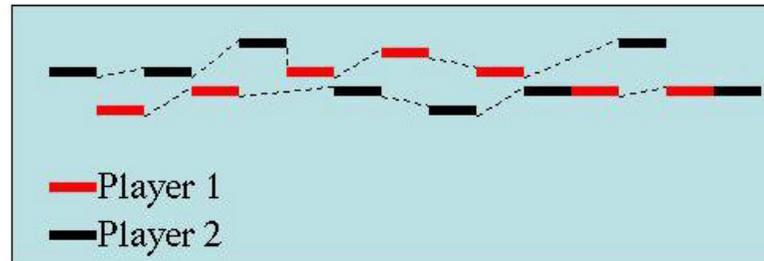


- Right-hand part segregates into two, isochronous streams.

7. Segregation of high tones from low tones in a Prelude by Bach
(Compound melody)

1. In Baroque music, one frequently finds instances where a single instrument plays a part that rapidly alternates between tones in different frequency ranges.
2. When this happens, the instrumental part is perceived to segregate into two concurrent ‘voices’ or melodic lines—that is, streams.
3. In music theory this is known as ‘compound melody’ or ‘virtual polyphony’.
4. Here’s quite a good example of this effect from the beginning of the Prelude in G major from Book 2 of Bach’s *Das Wohltemperirte Klavier*.
5. Note how the right hand part is perceived to segregate into two independent streams.
6. [PLAY BACH]

8. Streaming in African xylophone music (Bregman and Ahad, 1995, pp. 15–16, Track 7)

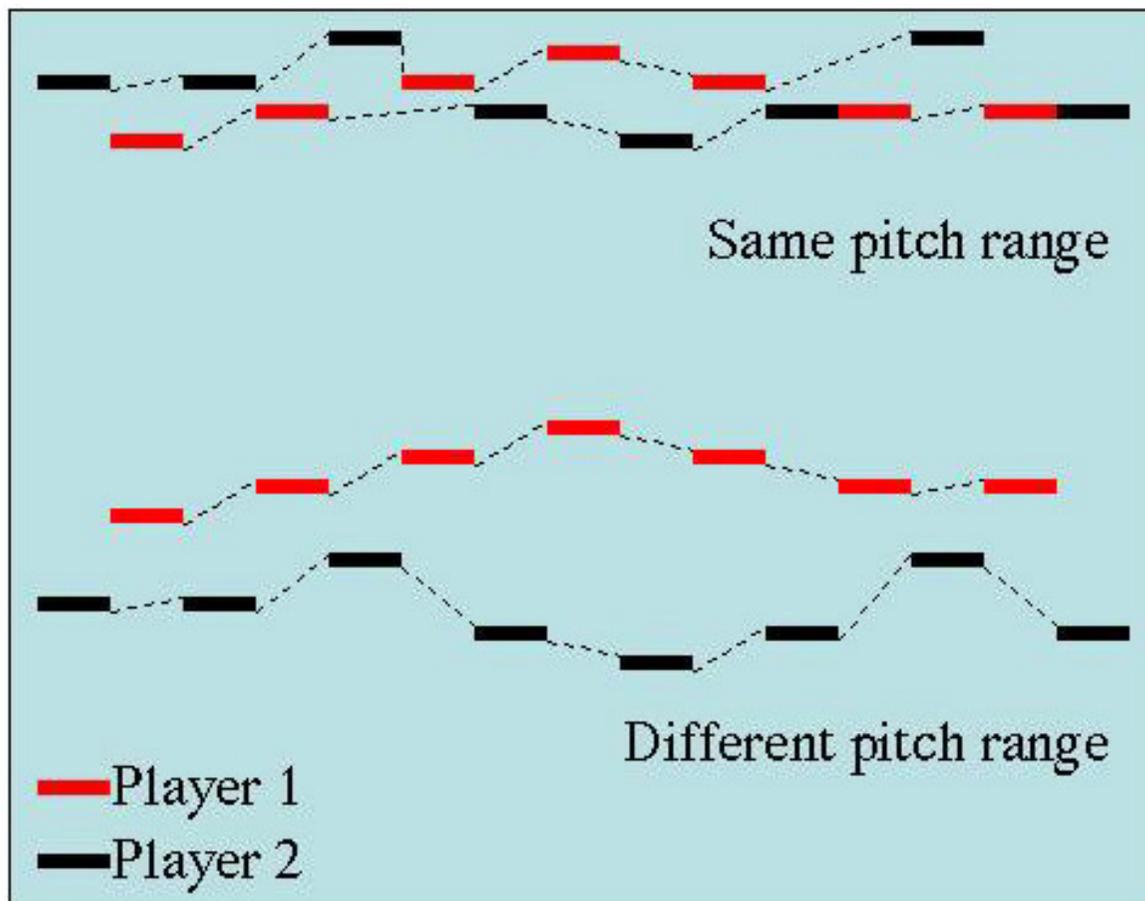


- Wegner (1990, 1993) identified some interesting instances of streaming in Ugandan amadinda music.
- Each of two players plays a repeating pattern, the notes of one player interleaved between those of the other.
- When heard separately, each part is isochronous.
- When heard together, combined sequence is segregated into two streams on the basis of pitch proximity giving rise to two streams with irregular rhythms.
- The perceived streams do not correspond to the individual parts played by the performers.

8. Streaming in African xylophone music
(Bregman and Ahad, 1995, pp. 15–16, Track 7)

1. Wegner (1990, 1993) has identified some interesting instances of sequential integration and stream segregation in Ugandan music written for a traditional type of xylophone called an amadinda.
2. Bregman and Ahad (1995, p. 17) explain that in this style of music, “each of two players plays a repeating cycle of notes, the notes of each player interleaved with those of the other.”
3. In other words, one player plays the odd-numbered notes, and the other plays the even-numbered notes.
4. The combined sequence is isochronous and each player’s part is also isochronous.
5. However, the combined sequence is perceived to segregate into two streams, one containing high pitches, the other containing low pitches.
6. And these two streams do *not* correspond to the two players’ isochronous parts.
7. Instead, the overall effect is two streams, each containing some notes from one part and some notes from the other, giving rise to an irregular rhythm in each of the perceived streams.
8. [PLAY TRACK 7]

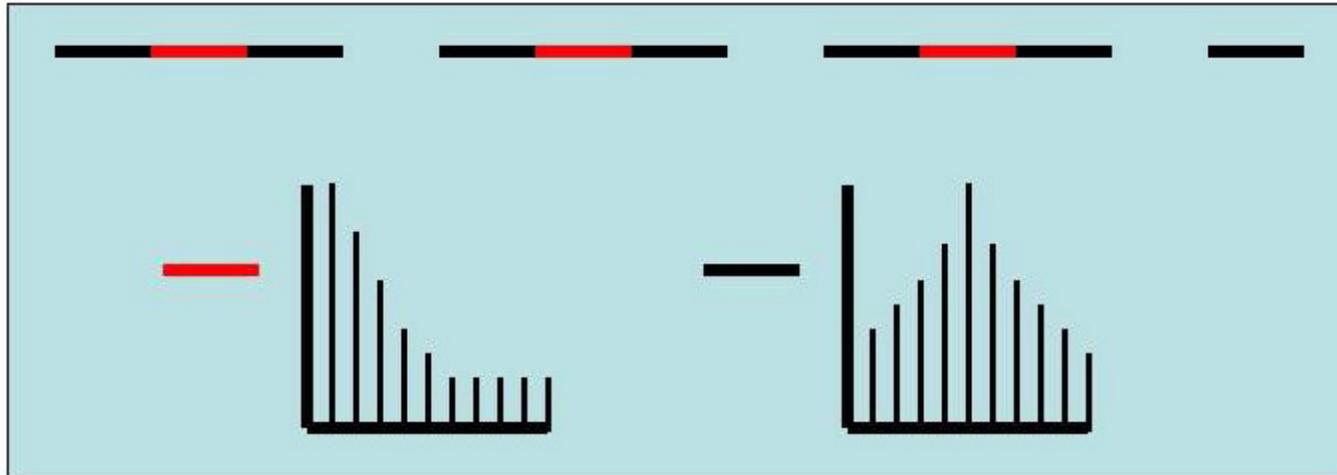
9. Segregating the two players' parts in Ugandan amadinda music
(Bregman and Ahad, 1995, p. 19)



9. Segregating the two players' parts in Ugandan amadinda music
(Bregman and Ahad, 1995, p. 19)

1. Recall that we were able to segregate a melody from interleaved distractor tones by transposing the melody so that its pitch range was sufficiently different from that of the distractor tones.
2. We can use the same idea to segregate one of the performer's parts from the other in the combined version of the amadinda music that we just heard in the previous demonstration.
3. In this example, one of the performer's parts has been transposed up an octave.
4. As with the previous example, the combined sequence again segregates into two streams. But this time, each stream corresponds to one of the performer's parts, whereas in the last example each stream contained notes from both parts.
5. This time, therefore, we perceive two streams and each one has an isochronous rhythm.
6. [PLAY TRACK 8]

10. Stream segregation based on timbre difference
(Bregman and Ahad, 1995, p. 21, Track 10)

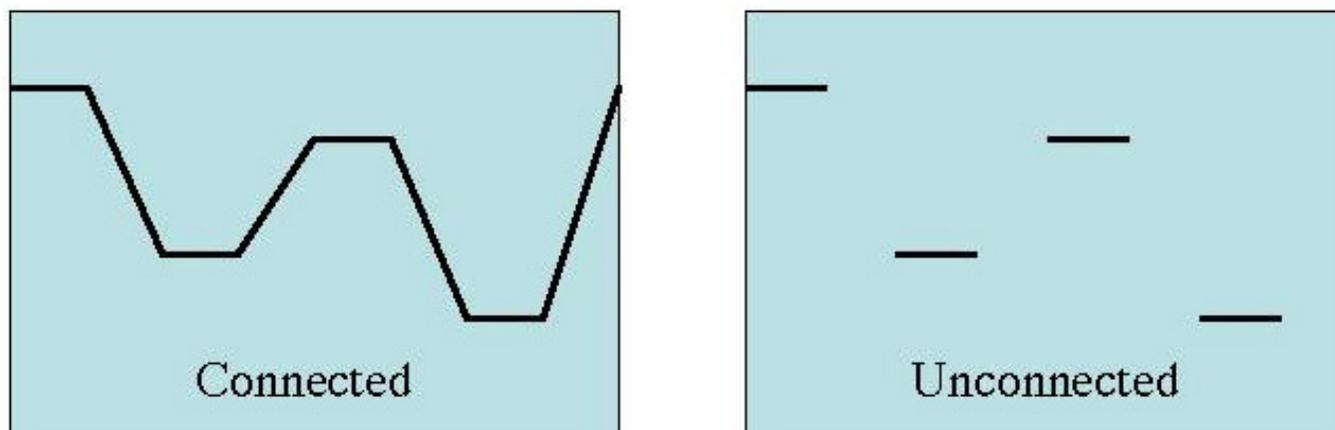


- If all three tones in the galloping pattern have the same fundamental but the middle tone has a different timbre from the first and third tones, then, at a particular speed of presentation, the sequence will split into two streams, each stream containing tones with a particular timbre.

10. Stream segregation based on timbre difference
(Bregman and Ahad, 1995, p. 21, Track 10)

1. So far we've shown that a sequence is segregated into two streams if the frequencies of the tones are clustered into separate frequency ranges and if the tones are presented rapidly enough.
2. Segregation of a sequence into separate streams can also be induced by using tones with different timbre: in general, we tend to assume that tones with similar timbre come from the same source and that tones with different timbre come from different sources.
3. Our auditory scene analysis system therefore tends to assign a separate stream to each set of tones in a sequence with a particular timbre.
4. In the following demonstration you'll hear the three-tone galloping pattern but in this case, all three tones in the pattern have the same fundamental frequency and therefore the same perceived pitch.
5. Clearly, if all three tones also had the same timbre, they would be integrated into a single stream even at a very high speed of presentation.
6. However, in the example I'm going to play you the middle tone in each three-tone pattern has a duller timbre than the first and third tones because its spectral peak is lower, as shown schematically in the diagram.
7. The result is that at a fairly moderate speed, the sequence is perceived to break up into two streams, one containing the tones with the brighter timbre and the other containing the tones with the duller timbre. This happens despite the fact that all the tones have the same fundamental frequency.
8. [PLAY TRACK 10]

11. Effects of connectedness on segregation (Bregman and Ahad, 1995, pp. 23–24, Track 12)

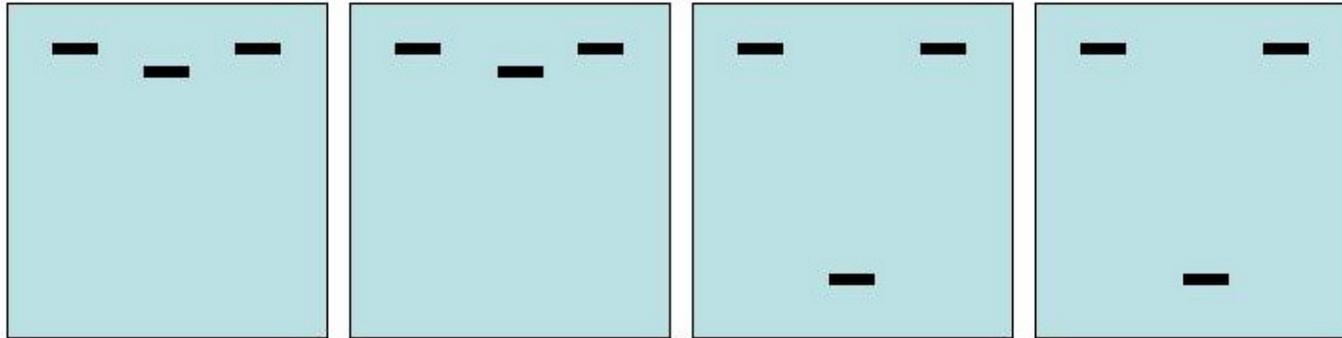


- Gestalt principle of good continuation also influences way we hear streams.
- Bregman and Dannenbring (1973) showed that inserting frequency glides between tones in a sequence promotes sequential integration.
- Could be just another manifestation of grouping by similarity or proximity.

11. Effects of connectedness on segregation
(Bregman and Ahad, 1995, pp. 23–24, Track 12)

1. We've seen that the Gestalt principle of proximity and similarity seem to play a part in determining the streams that we perceive in an auditory scene.
2. Another Gestalt principle, that of 'good continuation', also seems to influence the way we segregate a scene into streams.
3. The Gestalt principle of good continuation states that we prefer to group together elements in a scene that lie along smooth curves.
4. The effect of continuity or connectedness on streaming can be demonstrated by comparing the way that we perceive a sequence consisting of tones connected by frequency glides (see left-hand diagram) with one in which the glides are omitted.
5. In the next demonstration which is based on an experiment by Bregman and Dannenbring (1973), you're going to hear two sequences. Both sequences contain 20 repetitions of a 4-tone pattern consisting of two high tones interleaved with two low tones.
6. In the first sequence, the tones are connected together by continuous frequency glides as shown in the left-hand diagram. However, in the second sequence, the frequency glides are replaced by silences as shown in the right-hand diagram.
7. You should find that the tones in the connected sequence integrate into a single stream, whereas the unconnected sequence tends to segregate into two streams, one containing the high tones and the other containing the low tones.
8. [PLAY TRACK 12]
9. The demonstration shows that continuity between events promotes sequential integration.
10. Although, one could also interpret this as being just another example of events being grouped together by similarity of frequency or, equivalently, proximity within the frequency dimension.

12. Effects of stream segregation on timing judgements (Bregman and Ahad, 1995, pp. 25–26, Track 13)



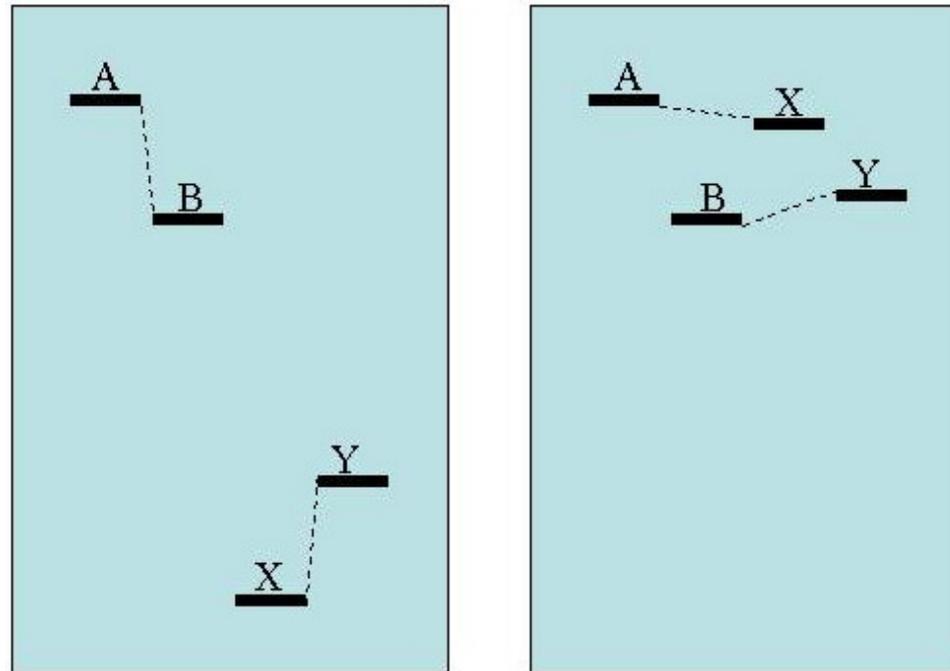
- It is much harder to perceive temporal relationships between tones when they are in different streams than when they are in the same stream.
- Demonstration based on experiment by van Noorden (1975).

12. Effects of stream segregation on timing judgements
(Bregman and Ahad, 1995, pp. 25–26, Track 13)

1. In the first demonstration, we noted that when the six-tone pattern was played quickly so that it segregated into two streams, it was difficult to detect that the tones in the lower-frequency stream occurred exactly between the tones in the high-frequency stream.
2. This exemplified the more general rule that it is difficult to estimate temporal relationships accurately between tones in different streams.
3. The demonstration you're about to hear, which is based on an experiment by van Noorden (1975), illustrates this result more directly.
4. This demonstration is in four parts.
5. In the first part, you'll hear an isochronous high tone. Then a second tone is added which is close in frequency to the high tone. Each of these lower tones is placed exactly mid-way between two of the high tones.
6. Because the lower tone is close in frequency to the higher tones, all the tones integrate into a single stream with a galloping pattern.
7. Because all the tones integrate into a single stream, it is easy to hear that each lower tone falls exactly mid-way between two higher tones.
8. In the second part of the demonstration, you will again first hear just the repeated higher tone on its own for a few cycles and then a lower tone, close in frequency to the higher tone is again added to form a galloping pattern.
9. However, this time, each lower tone is placed slightly after the mid-way point between two higher tones.
10. Again, because the higher and lower tones are close in frequency, they integrate into a single stream and it is easy to tell that each of the lower tones occurs slightly after the mid-way point between two of the higher tones.
11. The third and fourth parts of the demonstration both begin with a few repetitions of the high tone on its own and then a low, isochronous tone is added.
12. However, in the third and fourth parts, the frequency difference between the high and low tones is much larger than in the first two parts of the demonstration.
13. Your job is try to determine whether the low tones in the third and fourth parts of the demonstration fall exactly mid-way between the high tones as in the third diagram or slightly after the mid-way point between the high tones as in the fourth diagram.
14. When you listen to the third and fourth parts of the demonstration, write down whether you think the low tone falls exactly midway between the high tones or just after midway between.
15. [PLAY TRACK 13]

16. In the third and fourth parts of the demonstration the high and low tones segregate into separate streams because of the large frequency difference between them.
17. Consequently, it becomes very difficult to determine whether the low tones fall exactly midway between the high tones or slightly after the midway point.
18. Note that an analogous situation arises when you look at these diagrams: it is much easier to see that the low tone falls after the midway point in the second diagram than in the fourth diagram.
19. In fact, the low tone occurred after the midway point between the high tones in the third part of the demonstration and exactly midway between them in the fourth part.

13. Competition of frequency separations in the control of grouping
(Bregman and Ahad, 1995, pp. 28–29, Track 15)

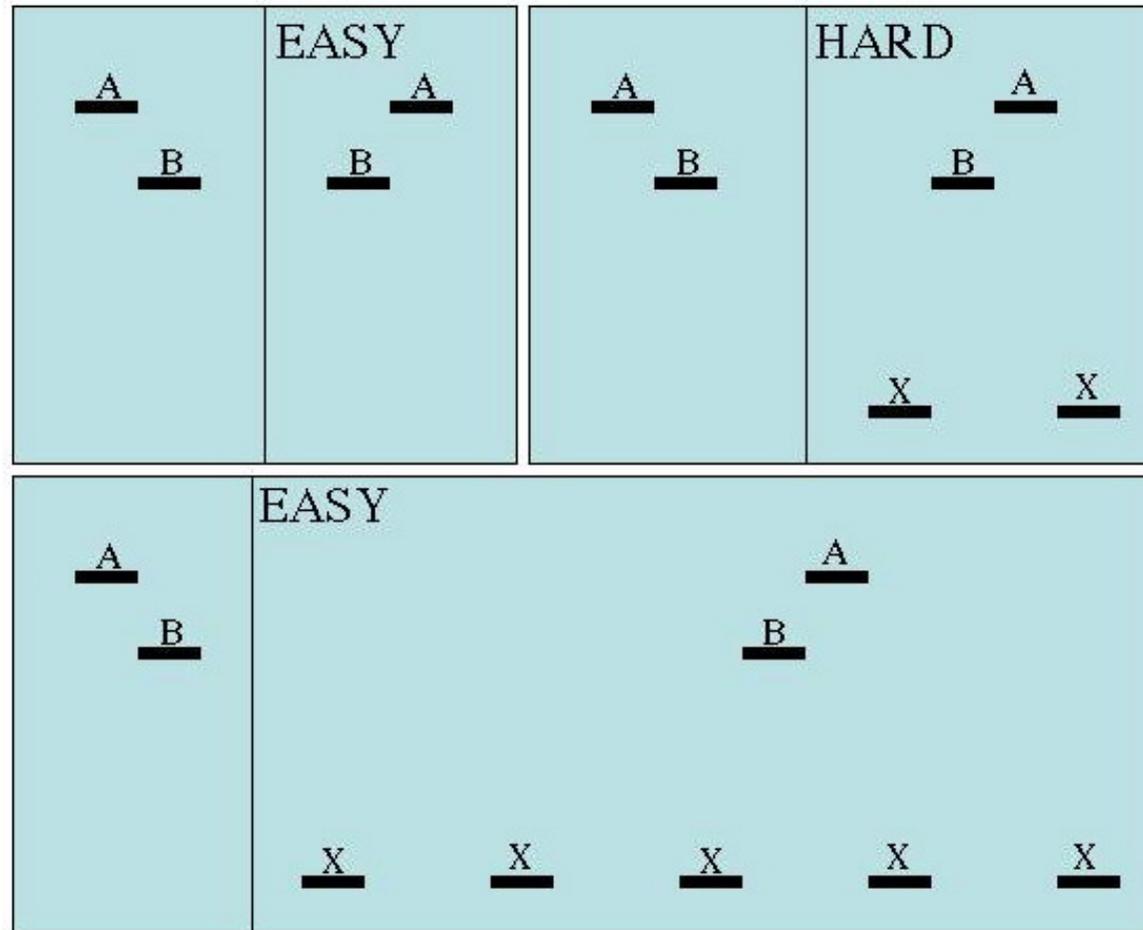


- When XY are far away from AB in frequency, it is easy to hear AB in ABXY.
- When X is integrated with A into the same stream and Y is integrated with B into a different stream, it is hard to hear AB in ABXY.

13. Competition of frequency separations in the control of grouping
(Bregman and Ahad, 1995, pp. 28–29, Track 15)

1. In this demonstration, based on an experiment by Bregman (1978), Bregman and Ahad (1995, pp. 28–29) show that a given pair of tones separated by some given time and frequency interval can be perceived as being either in the same stream as each other or in different streams, depending on the context in which they are presented.
2. We saw earlier that we are only able to detect a standard pattern when it is embedded in some larger pattern when all the tones in the standard pattern are in the same stream.
3. In the demonstration that I'm going to play you in a moment, you first hear a standard pattern consisting of two high tones AB forming a falling interval as shown in the diagram on the left.
4. You then have to try to continue hearing this two-tone standard pattern when it is presented with two more tones, XY, in a lower frequency range.
5. Because the tones XY are far away from AB in frequency, they integrate into a separate stream from A and B; and A and B integrate into the same stream as each other as shown in the diagram on the left.
6. In this case, it is easy to hear the AB standard pattern because both tones integrate into the same stream.
7. In the second part of the demonstration, the tones X and Y are raised in frequency so that X is close in frequency to A and Y is close in frequency to B. This causes A to be integrated with X into one stream and B to be integrated with Y into a different stream.
8. Because A and B are now in different streams, it is very hard to hear the standard pattern AB in the complete pattern ABXY.
9. [PLAY TRACK 15]

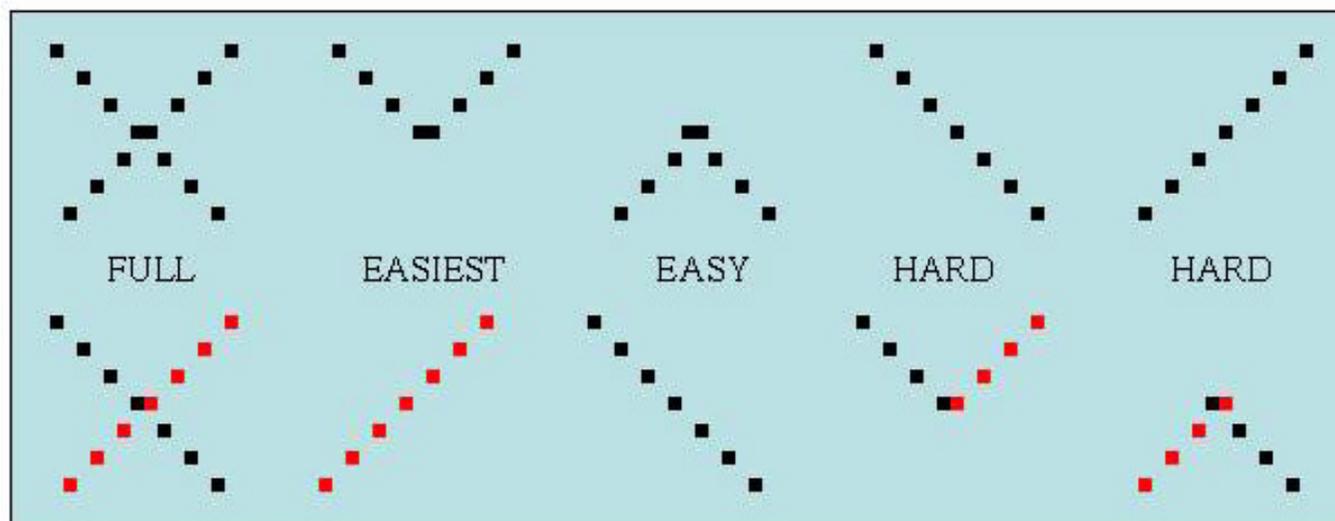
14. The release of a two-tone target by the capturing of interfering tones
(Bregman and Ahad, 1995, pp. 29–30, Track 16)



14. The release of a two-tone target by the capturing of interfering tones
(Bregman and Ahad, 1995, pp. 29–30, Track 16)

1. In an earlier demonstration we saw that when we segregate a sequence of tones into two streams, we do not do so as soon as the sequence starts but only after we have heard enough tones to accumulate enough evidence in support of a two-stream interpretation.
2. It is this principle of the cumulative effects of repetition that explains the way we perceive the demonstration that I'm going to play you in a moment.
3. In this demonstration, which is based on an experiment by Bregman and Rudnický (1975), you first hear a standard consisting of two tones, AB, followed by a two-tone comparison which is either AB or BA. When the pairs of tones are presented in isolation like this, it is fairly easy to tell whether the comparison is the same as the standard or different (see left-hand diagram).
4. Next, you hear a two-tone standard, AB, followed by a 4-tone comparison which consists of either AB or BA preceded and followed by a tone with a third frequency, X. Your task is again to determine whether or not the tones AB occur in the same order in the comparison as they do in the standard. The frequency difference between X and the tones A and B is quite large.
5. This time it is considerably harder to tell whether the two tones in the middle of the comparison are in the same order as they are in the standard.
6. Bregman (1990, p. 133) proposes that the task is more difficult in this second situation because, even though the tones X are in a different frequency range from the tones A and B, they do not form a separate stream from A and B because there hasn't been enough evidence accumulated to support a two-stream interpretation.
7. Instead, the pattern XABX (or XBAX) is interpreted as a structural unit and the first and last tones are heard as being more salient than the middle two tones precisely because they begin and end the pattern.
8. In the third part of the demonstration, you hear an isochronous sequence of tones X including the two tones that begin and end the XABX/XBAX unit.
9. The isochronous sequence of Xs preceding the occurrence of the segment containing the target tones A and B is long enough for our auditory scene analysis system to be induced into hearing the Xs as a stream.
10. The frequency difference between X and the target tones A and B is now large enough for A and B to be segregated from the X tones into their own separate stream.
11. Now, because A and B are in their own stream, separate from the X tones that precede and follow them, it is easier to hear whether or not the target is the same as the standard.
12. [PLAY TRACK 16]

15. The perception of 'X-patterns' (Bregman and Ahad, 1995, pp. 31–32, Track 17)



- Tougas and Bregman (1985) studied perception of X-patterns.
- Listeners hear a 'bouncing percept' unless forced to hear a crossing percept by emphasizing either the descending or the ascending sequence of tones by, for example, timbral differences between the sequences.

15. The perception of ‘X-patterns’
(Bregman and Ahad, 1995, pp. 31–32, Track 17)

1. Tougas and Bregman (1985) studied the way that we perceive *X-patterns*.
2. An X-pattern consists of two interleaved isochronous tone sequences, one ascending and the other descending as shown in the left-hand diagram.
3. Recall that we are very bad at hearing a standard pattern when it is embedded in a larger pattern in which it is not wholly contained within a single stream.
4. We can use this fact to find out how a listener segregates a given pattern into streams: if a listener can easily hear some given standard pattern within a larger pattern then that suggests that the standard is wholly contained within a stream in the larger pattern.
5. The principle of good continuation predicts that listeners will hear an X-pattern to be segregated into two streams, one consisting of the sequence of rising tones and the other consisting of the sequence of falling tones.
6. Such a percept is called a ‘crossing percept’.
7. However, it turns out that listeners find it very hard to hear the the complete sequence of ascending tones or the complete sequence of descending tones as part of the complete X-pattern.
8. This suggests that most listeners do *not* hear a crossing percept.
9. However, they find it very easy to hear a standard consisting of all the tones whose frequencies are greater than or equal to the middle tones (i.e., the V-shaped pattern in the second diagram on the top row) when this standard is embedded in the complete X-pattern.
10. They also find it easy to hear a standard consisting of all the tones whose frequencies are less than or equal to the middle tones (i.e., the inverted V-shaped pattern in the third diagram on the top row) when it’s embedded in the complete X-pattern.
11. This suggests that when listeners hear an X-pattern, they perceive the stimulus to be segregated into two streams, one containing the tones whose frequencies are higher than or equal to the middle tones (i.e., the upper V-shaped path in the second diagram on the top row), and the other containing the tones whose frequencies are lower than or equal to the middle tones (i.e., the lower, inverted-V-shaped path in the third diagram on the top row).
12. This way of hearing an X-pattern is called a ‘bouncing percept’.
13. However, it is possible to force listeners to hear a crossing percept by making the timbre of the tones in the ascending sequence different from that of the tones in the descending sequence.

14. If you do this, then it becomes easier to hear the ascending tone sequence and the descending sequence in the X-pattern than the two V-shaped patterns.
15. [PLAY TRACK 17]

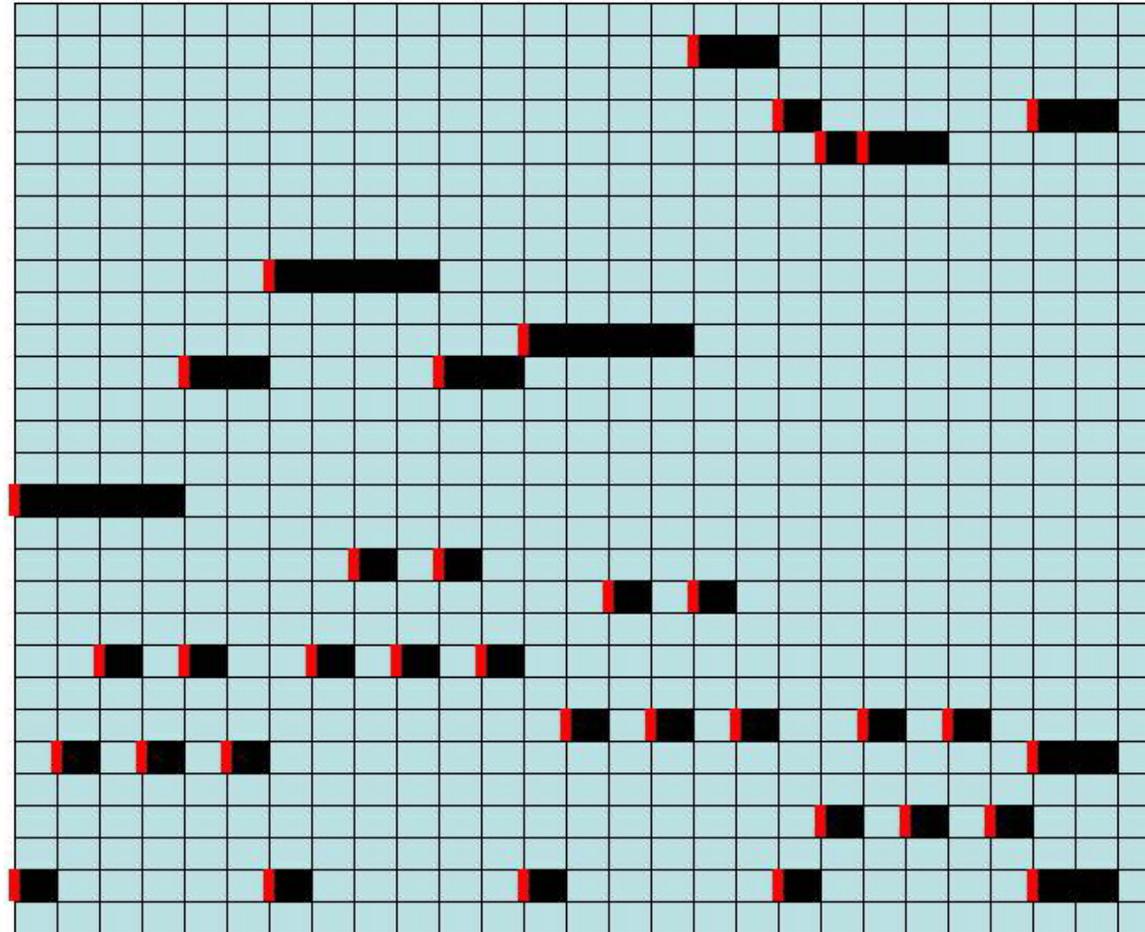
16. Temperley's (2001) Computational Theory of Music Cognition

- Theory is heavily influenced by GTTM.
- Presents models of metre, phrasing, counterpoint, harmony, key and pitch-spelling.
- Each model contains well-formedness rules and preference rules.
- Implemented as computer programs.
- Optimisation problem of finding best analysis solved using dynamic programming technique.

16. Temperley's (2001) Computational Theory of Music Cognition

1. I'm going to describe David Temperley's computational model of contrapuntal structure which is based on the work on auditory stream segregation that I've just been talking about.
2. Temperley's (2001) model of contrapuntal structure forms one part of the theory described in his book, *The Cognition of Basic Musical Structures*.
3. As I mentioned in the first lecture on metrical structure, in this book, Temperley presents a computational theory of music cognition that is deeply influenced by Lerdahl and Jackendoff's (1983) GTTM.
4. Like Lerdahl and Jackendoff, Temperley attempts to explain the cognition of common-practice music by means of a system that generates structural descriptions from musical surfaces'.
5. As in GTTM, the hypothesis underlying Temperley's theory is that the analysis it generates for a passage of music correctly describes certain aspects of how the passage is interpreted by listeners who are experienced in the idiom.
6. Like GTTM, Temperley's theory consists of a number of preference rule systems, each containing well-formedness rules that define a class of structural descriptions and preference rules that specify an optimal structural description for a given input.
7. Temperley presents preference rule systems for six aspects of musical structure: metre, phrasing, counterpoint, harmony, key and pitch spelling.
8. In collaboration with Daniel Sleator, Temperley has implemented most of his theory as computer programs.
9. As I've explained before, finding the analysis that best satisfies a set of preference rules is an example of an *optimisation problem* and one well-known technique in computer science for solving optimisation problems is the *dynamic programming* technique (Bellman, 1957; Cormen *et al.*, 1990, Chapter 16).
10. Each of Temperley's six preference rule models is implemented using the dynamic programming technique.

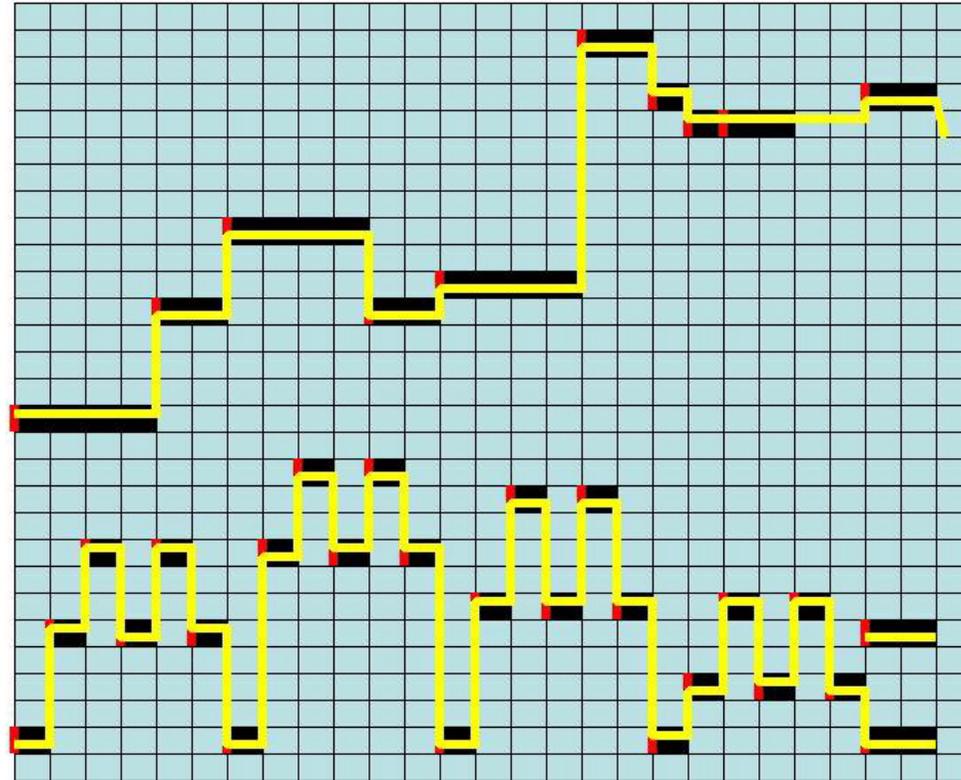
17. Temperley's (2001) Computational Model of Contrapuntal Structure: The Input Representation



17. Temperley's (2001) Computational Model of Contrapuntal Structure:
The Input Representation

1. The input to Temperley's (2001) model of contrapuntal structure must be in the form of a 'piano roll' that gives the onset time, duration and MIDI pitch number of each note.
2. Moreover, he assumes that the input is quantized to beats at the lowest metrical level in the passage.
3. As Temperley (2001, p. 96) points out, the fact that the input is quantized to the nearest chromatic pitch and the nearest metrical beat means that it can be represented as a two-dimensional grid of squares like the one shown here in the diagram.
4. Each of the squares in this grid is either white (actually blue, in this diagram), if there is no note there, or black if there is.
5. A red bar drawn along the left-hand edge of a square indicates a note onset. The offset of a note is indicated either by a white square or by the onset of another note with the same pitch (see example).
6. Note that this particular representation disallows overlapping notes of the same pitch.
7. Note also that because his input representation contains no information about timbre, loudness or spatial location, the streaming is done purely on the basis of temporal and pitch information.

18. Temperley's (2001) Computational Model of Contrapuntal Structure: The Well-Formedness Rules



CWFR 1 A stream must consist of a set of temporally contiguous squares on the plane.

CWFR 2 A stream may be only one square wide in the pitch dimension.

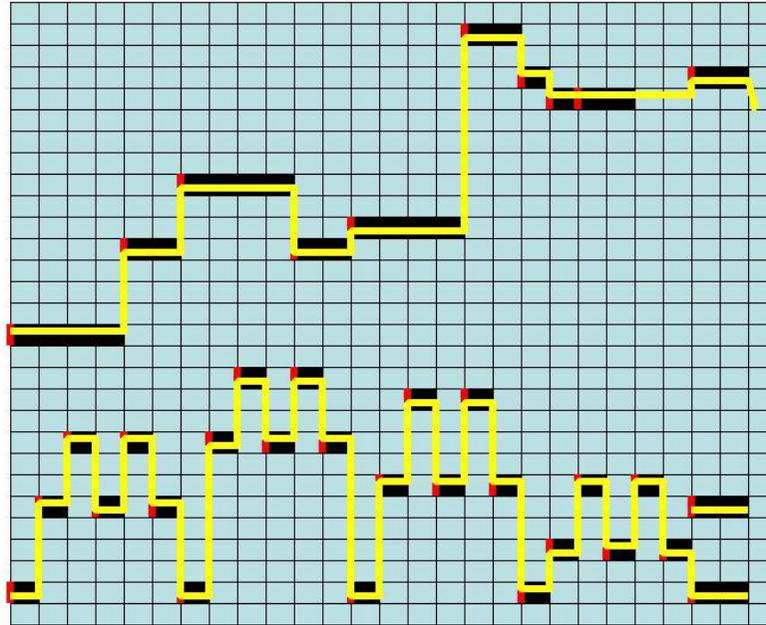
CWFR 3 Streams may not cross in pitch.

CWFR 4 Each note must be entirely included in a single stream.

18. Temperley's (2001) Computational Model of Contrapuntal Structure:
The Well-Formedness Rules

1. The output of Temperley's model of counterpoint is a description indicating the stream to which each note belongs.
2. This can be represented graphically as shown in the diagram where a continuous line is drawn through all the squares that belong to a particular stream.
3. Temperley's theory of counterpoint contains 4 well-formedness rules of which the first states that "A stream must consist of a set of temporally contiguous squares on the plane" (Temperley, 2001, p. 97).
4. This simply means that a stream cannot skip a column in the grid and then restart again. However, Temperley (2001, p. 98) stresses that a stream may be of any length—it does not have to span the whole piece.
5. CWFR 2 states that "a stream may be only one square wide in the pitch dimension" (Temperley, 2001, p. 98).
6. This implies that each object within a stream must be a single note. Now, in practice, it seems that the objects within streams in music need not be single notes—they may be chords, for example. But Temperley's theory does not take this into account: he assumes that each stream is a sequence of notes in which no two notes occur simultaneously.
7. CWFR 3 states that "streams may not cross in pitch" (Temperley, 2001, p. 98).
8. As we saw in the demonstration on the perception of X-patterns, listeners generally try to avoid crossing streams. This well-formedness rule reflects this experimental result.
9. However, streams do cross occasionally in music, suggesting that this well-formedness might have been better expressed as a very strong preference rule.
10. CWFR 4 simply states that "each note must be entirely included in a single stream" (Temperley, 2001, p. 99).

19. Temperley's (2001) Computational model of contrapuntal structure The Preference Rules



CPR 1 (Pitch Proximity Rule) Prefer to avoid large leaps within streams.

CPR 2 (New Stream Rule) Prefer to minimize the number of streams.

CPR 3 (White Square Rule) Prefer to minimize the number of white squares in streams.

CPR 4 (Collision Rule) Prefer to avoid cases where a single square is included in more than one stream.

19. Temperley's (2001) Computational model of contrapuntal structure
The Preference Rules

1. Temperley's (2001) computational model of counterpoint contains 4 preference rules.
2. CPR 1 states that we "prefer to avoid large leaps within streams" (Temperley, 2001, p. 100).
3. One of the fundamental results of the experimental studies in auditory scene analysis that I discussed earlier was that large differences in frequency promote stream segregation. CPR 1 is simply an expression of this principle.
4. CPR 2 states that we "prefer to minimize the number of streams" (Temperley, 2001, p. 101).
5. Recall that our auditory scene analysis system begins by assuming a one-stream interpretation of the scene and only switches to a two-stream interpretation after enough evidence for such an interpretation has been accumulated (e.g., by hearing many tones clustered into two frequency ranges).
6. CPR 3 states that we "prefer to minimize the number of white squares in streams" (Temperley, 2001, p. 101).
7. In music, notes frequently contain rests which would be represented as white squares in Temperley's input representation. This means that streams must be permitted to contain rests.
8. However, if the gap becomes too long, this clearly weakens the integration between events necessary for them to form a string.
9. Recall the demonstration I played you earlier on where we compared the perception of tones connected by frequency glides with the same tones separated by gaps.
10. CPR 4 states that "prefer to avoid cases where a single square is included in more than one stream" (Temperley, 2001, p. 101).
11. We only very rarely hear a tone as belonging to two different streams. One important example of such a phenomenon is in the bouncing percept of the X-pattern that I mentioned earlier on. If you remember, there the middle tones were heard as belonging to both the upper stream and the lower stream.

20. Temperley's (2001) Computational Model of Counterpoint

Testing the theory

- Comparison of predictions of rules with results of experiments on auditory stream segregation.
- Running program on Bach Fugues.

20. Temperley's (2001) Computational Model of Counterpoint
Testing the theory

1. I've already discussed on the previous slide some of the ways in which Temperley's preference rules relate to the results of studies on auditory stream segregation.
2. Temperley (2001, p. 108) also tested his theory by running the program that implements it on the first four fugues from Book 1 of Bach's *Das Wohltemperirte Klavier*.
3. One of the main differences between the output of Temperley's program and the contrapuntal structure of the fugues as indicated in the scores is that Temperley's program analysed the pieces into many more streams than there were voices in the fugues.
4. This is actually to be expected, however, since it is not uncommon for a voice to drop out of a fugue for several bars before re-entering. In such cases, Temperley's program would take the long gap to indicate that a stream had ended.
5. The program also tends to segregate compound melodies into two streams (e.g., the subject of the Fugue in C sharp major from Book 1).

References

- Bellman, R. (1957). *Dynamic Programming*. Princeton University Press.
- Bregman, A. S. (1978). Auditory streaming: Competition among alternative organizations. *Perception and Psychophysics*, **23**, 391–398.
- Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press, Cambridge, MA.
- Bregman, A. S. and Ahad, P. A. (1995). Demonstrations of auditory scene analysis: The perceptual organization of sound. Audio CD.
- Bregman, A. S. and Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, **89**, 244–49.
- Bregman, A. S. and Dannenbring, G. (1973). The effect of continuity on auditory stream segregation. *Perception and Psychophysics*, **13**, 308–312.
- Bregman, A. S. and Rudnicky, A. (1975). Auditory segregation: Stream or streams? *Journal of Experimental Psychology: Human Perception and Performance*, **1**, 263–267.
- Cormen, T. H., Leiserson, C. E., and Rivest, R. L. (1990). *Introduction to Algorithms*. M.I.T. Press, Cambridge, Mass.
- Lerdahl, F. and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. MIT Press, Cambridge, MA.
- Temperley, D. (2001). *The Cognition of Basic Musical Structures*. MIT Press, Cambridge, MA.
- Tougas, Y. and Bregman, A. S. (1985). The crossing of auditory streams. *Journal of Experimental Psychology: Human Perception and Performance*, **11**, 788–798.
- van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences*. Ph.D. thesis, Eindhoven University of Technology, Eindhoven, The Netherlands.
- van Noorden, L. P. A. S. (1977). Minimum differences of level and frequency for perceptual fission of tone sequences abab. *Journal of the Acoustical Society of America*, **61**, 1041–1045.
- Wegner, U. (1990). *Xylophonmusik aus Buganda (Ostafrika)*. Number 1 in Musikbogen: Wege zum Verständnis fremder Musikkulturen. Florian Noetzel Verlag, Wilhelmshaven. (Cassette and book).
- Wegner, U. (1993). Cognitive aspects of amadinda xylophone music from buganda: Inherent patterns reconsidered. *Ethnomusicology*, **37**, 201–241.