

# What Auditory Objects Are

In this paper I give a characterisation of auditory objects and argue that auditory object perception is governed by principles that mirror the principles that govern visual object perception.

The tenth variation of Bach's *Goldberg Variations* is a short four-voice fugue. It begins with a subject in the bass (on G below middle C) which is answered after four bars in the tenor; the soprano voice begins after a further four bars, and the final alto voice begins four bars later. When played at a normal tempo one doesn't hear the variation as a sequence of notes in the order in which they are played. As each new voice begins its notes are heard as a distinct melody – a separate voice parallel to the other voices – so that after twelve bars one hears four parallel melodies playing simultaneously. It is possible to attend selectively to each melody, but difficult to attend to more than one simultaneously; instead, when one attends to one melody it is heard against a background of the others. Although one can hear the order of and temporal relations between the notes that make up each melody, it is impossible to hear the order of the notes that make up the variation as a whole, and impossible to hear the temporal relations between the notes as they are actually played.<sup>1</sup>

This is an example of auditory grouping. We experience a sequence of notes as a number of separate groups or streams each of which is made up of only some notes in the sequence. This kind of grouping is central to our experience of many kinds of music, but grouping in music is the result of a process that occurs in, and is essential to, auditory perception generally – or so, at least, I shall argue.

## I. Which objects are auditory objects?

Sounds are those objects of experience that can be characterised in terms of their loudness, pitch, and timbre. We can regard any sound as an auditory object, but I am going to use the term 'auditory object' to mean those temporally extended sequences of sounds that are experienced *as* grouped in the way I have described. These

---

<sup>1</sup>It is impossible, at least, for someone who is not very familiar with the music.

sequences may or may not be composed of discrete, countable, elements; in what follows I will talk of elements of a sequence and mean either the discrete elements of a sequence or simply temporal parts of a sequence that are not experienced as discrete.<sup>2</sup> By reserving my use of the term auditory object to sequences experienced as grouped I don't mean to imply that there is some fundamental difference between auditory objects and other sounds. There is not. It is simply that the issues that I want to discuss arise most clearly for sequences.

Not all sequences of sounds are auditory objects (as I am using the term) because not all sequences of sounds are experienced *as* grouped. When we experience a sequence as grouped, we experience the sequence as a single, temporally extended, sound which may change through time; we experience different elements of the sound as belonging together. When we experience two or more sequences simultaneously (as we do with Bach's *Goldberg Variation*) we experience elements of each sequence belonging with other elements of the same sequence, but as not belonging with elements of the other sequence.

That elements are grouped together has consequences for our perceptual experience, in particular for our experience of properties of the group as a whole. For example, we experience melodic, rhythmic, or other patterns amongst elements of a grouped sequence but not amongst elements that belong to different groups; we can often tell the order of elements in a grouped sequence, but not the order of elements that belong to different groups;<sup>3</sup> and we can attend to one group to the exclusion of another, but not to two or more different groups simultaneously, nor can we simultaneously attend to two or more elements that belong to different groups. What I am calling auditory objects are those objects of auditory experience which have these characteristic features; they are what that I am going to discuss in this paper.

## II. Why do we experience auditory objects?

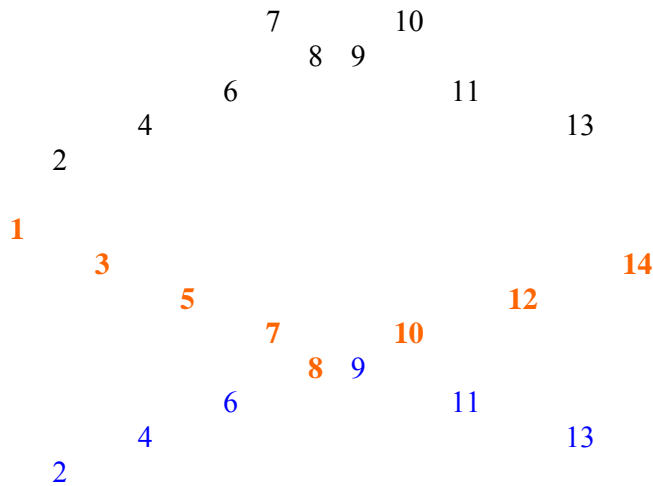
It is sometimes suggested that we can explain why we experience auditory grouping by appealing to principles of auditory experience analogous to Gestalt principles in

---

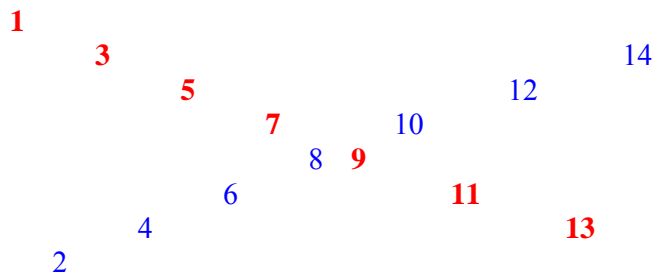
<sup>2</sup> Although different questions arise for the two kinds of sequence, as far as the issues that I discuss in this paper are concerned the differences don't matter.

<sup>3</sup> It's not always possible for listeners to perceive the order of elements in a group, but listeners can discriminate groups composed of different orderings of the same elements: the order of elements in a group determine how that group sounds; elements in other groups don't.





Pattern of tones heard as high and low pitched sequences



Pattern of tones heard as sequences of rising and falling tones crossing midpoint

In these examples, elements are experienced as organised into two groups and the groups are heard as distinct. Each element belongs to one group but not the other. What explains why the elements are grouped in the way they are? Altering the properties of the elements, or the speed at which they are played, changes the way they are grouped; it might seem, therefore, that grouping occurs according to some fairly simple principles. In particular, similar and proximal elements tend to be grouped together. There are different dimensions of similarity, and elements may be similar in virtue of their pitch or in virtue of their timbre with – as the second example demonstrates – timbre contributing more to similarity than pitch. So it seems that we explain why we experience the elements as grouped in terms of properties of the elements, such as their pitch, and the relations between elements, such as similarity or temporal proximity.

### III. What's wrong with purely auditory explanations?

The problem with explaining auditory grouping in purely auditory terms is that it although it tells us when or in what circumstances auditory elements are grouped, it doesn't tell us why grouping occurs in those circumstances; and it doesn't answer the deeper question of what is it to experience elements *as* belonging to a group or what it is to experience an element at one time as a continuation of an element at an earlier time. An auditory object is an experience of a sequence of auditory elements *as* grouped. What is it to experience elements *as* grouped?

The same problems apply to Gestalt explanations of visual grouping.<sup>6</sup> In vision, a matrix of dots can be seen as grouped either as four columns or as five rows (see diagram). A set of three visual elements is normally seen as two overlapping shapes (see diagram): we see the two elements x and y as grouped to form a single shape passing beneath z, and so see the three visual elements as two 'objects' one lying on top of the other. Similarly we normally see the two fragments in the diagram as parts of a single letter B (see diagram); we experience the elements as grouped into a single shape rather than as separate and unrelated elements. Just as in the auditory case, in vision distinct visual elements are experienced as grouped together into visual objects. Gestalt psychologists attempted to explain these examples of visual grouping in purely visual terms; in terms, that is, of the properties of the various elements that make up the visual image and their relations to each other. This kind of explanation can seem plausible because changing the arrangement and properties of visual elements changes how we experience them as grouped; by manipulating properties of the image – in particular the arrangement and appearance of elements – and noting how the elements of the image are experienced as grouped, it is apparently possible to discover the properties of the image that determine grouping.<sup>7</sup>

For example, whether we experience the matrix of dots grouped as four columns or as five rows depends on the spatial arrangement of the dots; we experience the dots as columns when the dots that make up the columns are closer to one another than to the other dots. This suggests that elements of an image that are spatially proximate tend to be seen as a group. In the case of one shape occluding

---

<sup>6</sup> As is often noted, for example by Bregman 1990, ch.2; Deutsch 1999; and Matthen 2005, p.287.

<sup>7</sup> This was roughly the approach of the Gestalt psychologists. They formulated various principles to explain visual grouping by appeal to properties such as spatial proximity, similarity, and good continuation. A particular experience of grouped elements is explained by showing how the elements satisfy the general principles of grouping. The principles explain grouping by appealing only to properties of the image.

another, we see element x as grouped with y only when there is a good continuation of their edges across the gap made by shape z; that is, their edges are such that they could be smoothly continued or joined under element z; when element x is offset relative to y disrupting their continuation, we no longer see them as grouped. In general, we experience an interrupted form as closed if the contour is ‘strong’ or ‘good’ at the point of interruption; that is, when the contours of the form continue smoothly on both sides of the interruption. This principle of closure explains why we see the letter fragments as parts of the letter B.

Explanations of grouping that appeal to Gestalt-type principles are unsatisfactory for two reasons. The first reason is that explaining why we experience a set of elements as grouped by showing that they satisfy a principle of grouping is uninformative. If we ask, for example, what the difference is between a set of elements that are grouped and a set of elements that are not grouped the answer is that the grouped elements satisfy a principle of grouping, i.e. that the grouped elements are, say, closer together or exhibit good continuity. That doesn’t tell us why we experience the elements as grouped, it just tell us that the elements are the same (in certain respects) as other elements we experience as grouped. We want to know why proximity or good continuation leads to grouping, why, in other words, the general principles are true.

The second reason is that appealing to such principles cannot explain what it is to experience a set of elements *as* grouped. When we experience a set *as* grouped each element is experienced as belonging together with other elements. What is the property of ‘belonging together’ that elements are experienced as having? There doesn’t seem to be an answer to that question in terms of the properties of the image that doesn’t simply appeal to the principles of grouping. That is, for a set of elements to be experienced *as* a group is for them all to be spatially proximate or for them to exhibit good continuity. Such an explanation is, again, uninformative: what it is to experience a spatially proximate set of elements as grouped is to experience them as spatially proximate.

We get a better explanation of visual grouping if we set grouping processes in the context of the function of the visual system as a whole. Visual grouping is the result of processes that operate as part of the processes that explain our capacity to see. The visual system functions to tell us about objects in our environment – what objects there are and where they are – in such a way that we can, for example,

recognise and act on those objects. It tells us about objects and surfaces on the basis of information extracted from the pattern of light reflected by them and detected by the retinas of the eyes. Because a typical visual scene comprises many different surfaces and objects, with closer objects partially obscuring those behind, many surfaces and surface regions of objects have no counterpart in the retinal image. Despite that, we don't experience surfaces as fragmentary, with the invisible regions of surfaces as non-existent; we experience occluded surfaces as continuing behind the objects and surfaces that occlude them. The visual system must, therefore, work out which visible parts of surfaces (parts that *do* have a counterpart in the retinal image) belong together as parts of single surfaces; that is, it must group together various parts of objects and surfaces that do have a counterpart in the retinal image according to whether they are parts of the same object. There must be a stage in visual processing that determines the layout of surfaces and objects in the environment; it produces a representation of the objects and surfaces that are most likely or that would best explain the parts of surfaces that it detects.<sup>8</sup> Visual grouping results from the operation of this stage of visual processing; we experience groupings as the result of the operation of the processes that enable us to perceive surfaces and objects in our environment.

The cues that determine how visual elements are grouped are cues that indicate the likely disposition of surfaces and objects in the environment of the perceiver. These cues may be relatively local properties of the image – T junctions, areas of equal luminance, and so on – but only because these local properties are evidence of objective surfaces. The two-dimensional images used by Gestalt psychologists produce visual stimuli that replicate some of these cues; they are treated by the visual system as if they were produced by surfaces and objects in a visual scene and grouped accordingly. Our experience of visual grouping can therefore be explained in terms of the processes that function to represent surfaces and objects.<sup>9</sup> In a similar way, we can explain auditory grouping in terms of the function of the auditory system.

---

<sup>8</sup> See Nakayama et al. 1995 for evidence for and a description of this stage of visual processing.

<sup>9</sup> Not only does that better explain the visual grouping examples that I described, it also enables us to explain examples of grouping describe by Nakayama et al. (1995) that cannot be explained in terms of image properties alone: e.g. the fish-head example, the fragmented Bs. In both cases, image properties and relationships amongst elements are the same, but altering depth cues produces a different grouping and a representation of different surfaces.

#### IV. Auditory Function

What is the function of auditory perception? Auditory perception, like vision, functions to tell us about objects in our environment; it does so by detecting disturbances in the air produced by those objects, and by events involving them. Although it is perhaps obvious that vision functions to tell us about objects, that auditory perception does so too is not obvious. I shall begin by outlining an account of auditory function that justifies that claim, and then go on to say something about the consequences of that for understanding auditory grouping and auditory objects.<sup>10</sup>

Imagine that you are woken up in the middle of the night by a strange sound. As you lie there, listening, you can attend to your experience in two ways: you might attend to the sound itself, focussing on its attributes – its pitch, timbre, and loudness – but it is more likely that you will attend to what is making the sound: that it is the sound of a window breaking, that it came from the kitchen, and that now you can hear the sash being opened. When people are asked to describe what they hear (in psychoacoustics experiments, for example) they are often encouraged to attend to their experience in the first way: to describe the sensory attributes of the sounds they hear in abstraction from whatever it was that produced the sounds.<sup>11</sup> They may be helped by being played harmonically simple sounds produced by a tone generator, sounds which develop little over time and which have little or no ecological significance. There is little to describe about such an experience over and above the sensory qualities of the sounds. The majority of the sounds we hear are not like that, and most everyday listening is of the second kind: we attend to the apparent sources of the sounds we hear and listen to the things going on around us, to the objects and events that produced the sounds. In most everyday listening we are concerned with properties and attributes of the sound producing events and the environment in which they occur, rather than with properties of the sound itself.<sup>12</sup>

---

<sup>10</sup> The account in this section draws on the more detailed discussion and defence given in my . . . .

<sup>11</sup> When they do this, listeners adopt what Gaver (1993) calls a ‘musical’ and Scruton (1987, pp.2 ff.) an ‘acousmatic’ attitude to what they hear.

<sup>12</sup> In what follows I am going to talk about the perception of what may be labelled ‘ecological’ sounds and their sources – those sounds produced by naturally occurring events or various kinds. Ecological sounds themselves, rather than their sources, are not very interesting or informative; indeed for many ecological sounds it is actually rather difficult to attend to the sound rather than to the source of the sound and we are poor at describing the character of the sound. It is easy to overlook the features of ecological sound perception because we have become so used to hearing artificially produced sounds.

Although relatively little investigation has been done to determine how good we are at perceiving and recognising sound sources, that which has been done has found that we are surprisingly good at both. There is evidence that we are capable of recognising very specific characteristics of the events and objects we hear.<sup>13</sup> We are, for example, very good at recognising what kind of object or event produced a sound. Listeners who were played recordings of different size jars and bottles falling to the ground and either bouncing or breaking and were asked which kind of event – a bouncing or a breaking – they heard were almost always correct.<sup>14</sup> When asked to identify thirty common natural sounds in a free identification task – sounds such as those produced by clapping, tearing paper, and footsteps – listeners recognised source events very reliably; they described the sounds in terms of the objects and events which caused them, and only described the sensory qualities of sounds whose source events they could not recognise.<sup>15</sup> In a similar experiment in which seventeen sounds were played, listeners were asked to identify what they heard. They nearly always described the sounds in terms of their sources, and were surprisingly accurate. Several participants could readily distinguish the sounds made by someone running upstairs from those of someone running downstairs; others were correct about the size of objects dropped into water; and most could tell from the sound of pouring liquid that a cup was being filled. Some sounds – such as the sound of a file drawer being opened and closed – were difficult to identify, but the listeners’ descriptions revealed what might be regarded as basic attributes of what was heard: “several people said the file drawer sounded like a bowling alley, both of which might be described as ‘rolling followed by impact(s)’”.<sup>16</sup> Further investigation is likely to reveal many more examples.

---

<sup>13</sup> For a recent survey of much of this evidence, see Carello et al. (2005). Compare what I say here to accounts of visual object recognition, which has been studied in great detail and is widely understood to be a perceptual phenomenon with the results of the process of object recognition entering into the content of visual experience. I know of few studies of auditory object recognition, but see McAdams (1993) and Peretz (1993).

<sup>14</sup> Listeners’ success rate was 99%; see Warren and Verbrugge, 1984.

<sup>15</sup> The success rate was about 95%; see VanDerveer, 1979.

<sup>16</sup> Gaver, 1993a, p.12. It is plausible to suppose that recognising such events involves the perception of simpler, more fundamental, properties of events and that such properties may be perceived even when the event is not recognised. In much the same way visual recognition of an object as, for example, a television, involves perceiving the object as having more fundamental properties such as size and shape which it may be perceived as having even when it is not recognised as a television.

As well as recognising the sources of sounds we can perceive their properties. We are, for example, able to perceive the trajectory of an approaching sound source,<sup>17</sup> and the time to contact – that is, the time at which we will collide – with a sound source that is moving towards us.<sup>18</sup> We are good at hearing whether an invisible object making a noise is within reach;<sup>19</sup> and we are able to hear just as well as we can see whether a gap between a sound source and a vertical surface is wide enough to pass through.<sup>20</sup> We can identify the material composition of an object from the sound of an impact,<sup>21</sup> and perceive the force of the impact.<sup>22</sup> More surprisingly, perhaps, we are able to distinguish geometrical properties of objects. When differently shaped – circular, square, and triangular – flat steel plates of the same mass and surface area were suspended and struck by a steel pendulum released from a fixed location, listeners sitting behind a screen were able to classify the shapes at a level well above chance. A similar experiment was conducted with rectangular steel plates of different proportions and dimensions chosen so that all were equal in mass and surface area. Listeners had to respond by adjusting lines to provide a visual match for the height and width of the plate. Although they were given no other information about the size of the object, the actual linear dimensions of the plates accounted for 98% of the variance in the listeners’ responses.<sup>23</sup> Similarly, when listeners were asked to indicate the lengths of cylindrical rods dropped to the floor, the actual length of the rods accounted for 95% of the variance in perceived length.<sup>24</sup>

Given that we can perceive and recognise the sources of the sounds we hear, it is plausible that auditory perception functions to tell us about those sources, and that in addition to representing sounds, our auditory experience represents the sources of sounds and their properties. But how can auditory perception have this function? To understand how, we need to understand how sound sources produce sounds. Sounds can be produced by many different kinds of things – liquids, solid objects, strings, air

---

<sup>17</sup> Neuhoff, 2004.

<sup>18</sup> Schiff and Oldak, 1990.

<sup>19</sup> Carello et al., 1998.

<sup>20</sup> Russell and Turvey, 1999.

<sup>21</sup> Wildes and Richards, 1988.

<sup>22</sup> Freed, 1990.

<sup>23</sup> The plates were a square (482mm), a medium rectangle (381mm x 610mm), and a long rectangle (254mm x 914mm), the width indicator ranged from 0 to 2.5m, and the height indicator from 0 to 1.5m. Although listeners’ relative scaling of the plates was accurate, the perceived dimensions were underestimates of actual dimensions, ranging from 252mm to 445mm for an actual range of 254mm to 914mm (Kunkler-Peck and Turvey 2000).

<sup>24</sup> Carello, et al., 1998.

movement – and in many different ways, but for simplicity I am going to consider only material objects. Material objects produce sounds when they are struck, tapped, scraped, broken or otherwise caused to vibrate.

We often picture a vibration as a single sine wave. Not even something as simple as a plucked string vibrates in such a simple way. The vibration of a plucked string is complex; it comprises a number of simple vibrations at frequencies which are integer multiples of the lowest, or fundamental, frequency of the vibration.<sup>25</sup> Any complex vibration is equivalent to a number of simple frequency components superimposed on each other; that means we can represent any complex vibration as a pattern or structure of individual frequency components. Objects vibrate along a greater number of dimensions than strings and consequently their vibrations are more complex and so composed of a greater number of frequency components. What's important for our understanding of auditory perception is that the particular pattern of frequency components produced by a material object when it vibrates is determined in a law-like way by both the physical nature of the object and the nature of the event that caused it to vibrate. For example, the shape and size of the object determine the lowest frequency of its vibration and what harmonics are present. The overall amplitude of the vibration is determined by the force that initially deforms the object, but because objects are not linearly elastic the amplitude of individual frequency components varies with the force of the initial deformation. The spectral composition of the vibration therefore changes according to how hard the object was struck.<sup>26</sup> Vibrating objects lose energy over time and their vibration decays. The rate of decay of different frequencies components – and so changes in the spectral composition of the vibration over time – is determined by the material of which the object is composed.

The pattern of frequency components that comprise the vibration of an object and the way that pattern changes over time is determined by the nature of the object and the nature of the events that caused it to vibrate. That pattern and the way it changes therefore embody information about the object that produced the vibration and the event that caused it to vibrate. The vibrations of objects produce compression waves in the surrounding air. In an enclosed space, the compression waves will

---

<sup>25</sup> The vibration of a plucked string is made up of the odd harmonics of the fundamental unlike the vibration of a string excited in some other way, which includes both odd and even harmonics.

<sup>26</sup> It is in virtue of this that we can distinguish the intensity of a sound (its loudness) from the apparent force of the impact that produced the sound.

reflect off surfaces and objects, and the waves produced by different objects will interact with each other to alter the spectral composition of the wave in determinate ways, with the result that the local disturbance of the air at any place will carry information about any number of objects and events, and about the environment in which they occur. This local disturbance of the air is what is detected by our ears.

The auditory system detects the frequency components that make up the complex vibrations of the soundwave that reaches the ears. To tell us about the sources of sounds it must construct a representation of objects producing sounds by extracting the information about them embodied in the pattern of frequency components detected by the ears.

If the sounds we heard were only ever produced by one object at a time the fact that a soundwave is made up of many frequency components would be unproblematic: components that are detected simultaneously would have been simultaneously produced by a single event, and successively detected components would have been produced by temporally successive parts of that event. Often, however, there are many different objects producing sounds simultaneously, so the compression wave that is detected by the ears is, at any moment, the result of the additive combination of the compression waves produced by all the sound producing events occurring in our immediate environment; as a result this compression wave is composed of frequency components produced by different objects.

Auditory perception therefore requires perceptual processing much like that involved in visual perception. We can think of the frequency components detected by the ears as analogous to the pattern of light detected by the retinas of the eyes. Just as we see things in virtue of detecting a pattern of light on a surface (the retina), so we hear things in virtue of detecting properties of soundwaves disturbing a surface (the basilar membrane). We don't, of course, see the pattern of light: our visual experience is the result of perceptual processes to which the pattern of light detected by the retina is one of the inputs. Similarly, we don't hear the frequency components of soundwaves detected by the ears; our auditory experience, including the sounds we hear, is a result of perceptual processes, to which the frequency components of soundwaves are one of the inputs. This perceptual process involves at least the following three stages.

The first stage is sensory transduction or detection: the ears detect properties of the soundwave – the local disturbance of the air. The result of this sensory

transduction is, in effect, a temporal spectrogram of the soundwave which encodes the frequency and temporal properties of the soundwave's vibration. The ears detect each of the frequency components (within a detectable range) present in the soundwave's vibration.

The second stage involves grouping together the frequency components that have been produced by the same source. Information about objects and events is embodied in the relationships amongst the frequency components produced by an object's vibration and the way those frequency components change over time. In order both to determine how many objects are producing sounds at any time and to extract information about those objects the auditory system must organise frequency components into groups corresponding to the objects that produced them. Frequency components need to be grouped so that all the frequency components produced by a single source are treated together, and those from different sources treated as distinct by subsequent processes. There are two kinds of grouping.

Firstly, frequency components produced at a time must be grouped together as having been produced simultaneously by a source; secondly, simultaneous groups must be sequentially grouped over time as having been produced by a temporally extended event involving a single object, and series of such sequences grouped as having been produced by a series of events involving a single object. In order, for example, for the auditory system to determine how many objects are producing sounds at any time, it must group the frequency components it detects at a time according to the object that produced them. If the auditory system detects sequences of frequency components then grouping them together as having been produced by a single object allows information about that object to be recovered: information about how the object is changing or moving, for example; and it allows events to be recognised. In order to recognise an object as dropped onto a hard surface and bouncing, for example, the auditory system must group the sequences of frequency components produced by the object as having been produced by a single object; similarly, in order to recognise water filling a glass we must experience a single continuous sound – the auditory system must group earlier and later frequency components as parts of the same group – as produced by an ongoing process. In both cases, in order to perceive the sources of sounds and their properties the auditory system must simultaneously and sequentially group frequency components.

How does the auditory system determine which frequency components to group together? In the case of simultaneous grouping, there are relationships that exist between components produced by the same source that are unlikely to exist between components produced by different sources. For example, an object's vibration often has frequency components that are harmonics of a fundamental frequency and so the frequency components of a soundwave that are produced by the same source will often be harmonically related. Such harmonic relationships are unlikely to exist between frequency components produced by distinct sources since it is unlikely that two simultaneously occurring natural events produce overlapping sets of harmonics. This means that if the auditory system detects a number of frequency components that are harmonically related then they are likely to have been produced by the same source. Similarly, the soundwave produced by a single event will have frequency components that share temporal properties – all the components will begin at the same time – are likely to be in phase with one another, and are likely to change over time in both amplitude and frequency in similar ways. Components produced by distinct sources are very unlikely to be related to each other in these ways. When the auditory system detects these relationships between components it groups them together and treats them as having been produced by the same source. Components that are not related in this way are not grouped together.

Just as for simultaneous grouping, there are relationships between frequency components produced by the same source at different times that are unlikely to exist between components that are produced by different sources. For example, sets of components with the same spectral composition at different times are unlikely to have been produced by different sources. Sources can change in size and so the frequency components they produce can shift in frequency. When this happens the overall pattern of frequencies is likely to remain the same. If the auditory system detects frequency modulated sets of components with the same spectral composition then they are likely to have been produced by an object that is changing. Objects cannot change instantaneously, so if successive frequency components differ greatly in frequency they are likely to have been produced by distinct sources. If two identical sets of components separated by a gap are detected, then it is more likely that they are produced by a single object than by two different objects. These examples are of bottom-up or stimulus driven grouping. It is likely that some grouping of sequences of components is top-down. That is, some sequences are grouped because they fit

into a perceptual pattern that the auditory system recognises as likely to have been produced by a certain kind of source.<sup>27</sup> Grouping sequences of familiar mechanical sounds is likely to be the result of such top-down grouping.

In general the auditory system makes best sense of the frequency components it detects, where making best sense of frequency components means grouping them – both simultaneously and sequentially – in such a way that they correspond to the sources that would best explain their occurrence. It is an important consequence of this that we cannot explain why the auditory system groups the frequency components that it detects in the way it does other than in terms of a process that functions to extract information about the objects that produced those frequency components. This is true of both simultaneous and sequential grouping. The auditory system groups together all and only frequency components that are likely to have been produced by the same source.

The third stage of processing involves extracting information from the frequency component groupings. The grouping process results in sets of frequency components that are treated by subsequent processes as having been produced by a single source. These sets of components carry information about those sources, and the fact that we can perceive various properties of the sources of sounds means that the auditory system must extract that information. Exactly what information is extracted and how it is extracted is still, for the most part, unclear. We can perceive how many sources there are, and often where they are; we can perceive various features of sources; and are able to recognise sources as events of certain kinds or involving certain kinds of object. The information extracted must be sufficient to explain these capacities. Recognition processes might match representations of the features of sources with representations of kinds of events and objects, or they might simply track some characteristic pattern of frequency components produced by certain kinds of events and objects. However exactly the information extraction and object recognition processes work, we know that they must be sufficient to explain our capacity to perceive and recognise the sources of sounds.

I have characterised the psychological processes involved in auditory perception; how do these processes relate to our auditory experiences and in particular our experiences of sounds and auditory objects? In virtue of their operation we

---

<sup>27</sup> Bregman calls this ‘schema-based organisation’: it involves ‘the activation of stored knowledge of familiar patterns or schemas in the acoustic environment’ (Bregman, 1990, p. 397).

perceive both sounds and their sources. What sounds we experience and how we experience them to be is determined by the way the auditory system simultaneously groups the frequency components it detects: the sounds we hear correspond to simultaneous frequency component groupings. If the auditory system groups the components it detects into a single group then we experience a single sound; if it groups them into two groups, then we experience two sounds. Given that the auditory system groups frequency components that are likely to have been produced by the same source, the sounds we experience normally correspond to their sources – to the things that produced them. How we experience sounds as grouped is determined by the way the auditory system sequentially groups frequency components. If the auditory system sequentially groups a sequence of frequency components then we experience the corresponding sounds as grouped; we experience a sequence of sounds as grouped in virtue of the auditory system having sequentially grouped the corresponding sets of frequency components. What sounds and auditory objects we experience is determined by the way the auditory system groups the frequency components it detects; we can only explain why the auditory system groups frequency components as it does in terms of a process that functions to tell us about the sources of those sounds and auditory objects; therefore, we can only explain why we experience the sounds and auditory objects we do in terms of a process that functions to tell us about their sources.<sup>28</sup>

On this view of auditory perception, sounds just are a certain pattern or structure of frequency components, and an experience of a sound represents a pattern or structure of frequency components instantiated by the soundwave that is detected by the ears. An experience of a sound is veridical just in case it is produced by the pattern or structure of frequency components that would normally produce that experience; it is not veridical if it is not produced by any such pattern or if it is produced by a pattern that would not normally produce that experience.

The auditory system functions to represent sounds and auditory objects that correspond to their sources (to the objects and events that produced them) as part of a process that extracts information about those sources. As a result our experience

---

<sup>28</sup> There is not space to defend this in detail, but it's worth noting how my view contrasts with an alternative. According to the alternative sound sources produce sounds; auditory perception functions to perceive sounds; we can tell things about the world on the basis of perceiving sounds in virtue of our knowing what causes certain kinds of sounds. Given the way auditory perception works, this alternative cannot be right. For a more detailed discussion, see my...

represents both sounds and the sources of sounds, and we normally experience sounds that correspond to their sources.

## **V. Explaining auditory grouping**

This account of the function of auditory perception provides an explanation of our experience sounds as grouped – an explanation of why we experience auditory objects. We experience sounds as grouped as the result of a process that functions to tell us about the sources of sounds. Sequential grouping is a necessary step in the process of extracting the information about sound sources – in particular, about events involving them – that enables us to perceive and recognise them.

How does this explanation in terms of function relate to the explanation in terms of the principles of grouping? To the extent that the principles of grouping are true generalisations about experience, the explanation in terms of the function tells us why those principles hold. They hold because sounds grouped in accordance with the principles are likely to be grouped in a way that corresponds to their sources. For example, the pitch a sound is experienced to have is determined by the fundamental frequency of the object vibration that produced it.<sup>29</sup> The fundamental frequency of an object's vibration is determined by the size and shape of the object. It is unlikely that two identical naturally occurring objects produce sounds in the same period of time; therefore, a sequence of sounds all of which have the same pitch are likely to have been produced by a single object. The pitch of sounds produced by an object can change only if, and as fast as, the object changes. A sequence of sounds all of which have closely related pitches may have been produced by a single object that is changing in size; but two sounds with very different pitches are unlikely to have been produced by the same object. Sounds grouped according to the proximity of their pitches will, therefore, be grouped in a way that is likely to correspond to their sources. Most naturally occurring sounds do not have a pure pitch; they have timbre determined by the complex spectral composition of frequency components that determine them. The vibrations of two naturally occurring objects are very unlikely to have exactly the same spectral composition; therefore, a sequence of sounds all of which have the same spectral composition or timbre are very likely to have been

---

<sup>29</sup> This is somewhat simplified, but not in a way that affects the argument.

produced by the same object and two sounds with different timbres are very unlikely to have been produced by the same object. Sounds grouped according to their timbres will, therefore, be grouped in a way that is likely to correspond to their sources.

Explanations of grouping in terms of the principles of grouping are consistent with explanations in terms of the function of auditory perception, but explanations in terms of function are better explanations because they tell us *why* the principles are true. Furthermore, although the principles of grouping may be consistent with explanations in terms of function, and may describe generalisations that are true of experience, it doesn't follow that the auditory system uses or follows the principles in determining how to group frequency components. It is unlikely that the auditory system groups sounds *because* they are similar in pitch or similar in timbre – it is unlikely that those sensory qualities of sounds are causally explanatory of grouping. It is more likely that it groups sets of frequency components because they have the same spectral composition, and that those sets of frequency components with the same spectral composition determine experiences of sounds with the same timbre. The *consequence* would then be that similar sounds are grouped together as described by the principles; thus the truth of the principles emerges as a consequence of the way the auditory system functions, but is not explanatory of its function.

## **VI. What are auditory objects?**

An account of auditory grouping will be incomplete if it cannot give some account of what it is to experience a sequence of elements *as* grouped. Appealing to the function of auditory perception allows us to provide such an explanation.

I have argued that auditory experience represents both sounds and the sources of sounds. How do we experience the connection between the sounds we experience and the sources of those sounds? We don't experience the source of a sound independently of experiencing the sound that it produces. When we experience a sound we experience it as apparently having been produced a source of a certain kind. For example, in experiencing the sound produced by a solid object falling onto a hard surface we experience a sound as apparently having been produced by a solid object falling onto a hard surface; in experiencing the sound made by a bird singing outside the window we experience a sound as apparently coming from outside. Normally, when we hear a sound we hear it as having been produced by a source; in virtue of

that we can hear the source. That we hear sounds as produced by their sources is reflected in the way we describe sounds: we talk of the sound *of* a dropped ball and *of* a bird singing. Describing a sound as the sound *of* something can be naturally understood to mean the sound *made by* or *produced by* that thing.

Sounds are produced by sources: a sound has the property of having been produced by a source of a certain kind. When we experience a sound as having been produced by a source, our experience represents it as having that non-intrinsic property. Therefore, our auditory experience represents sounds and the sources of sounds and it represents sources *as* the sources of sounds by representing sounds as having a non-intrinsic property – the property of having been produced by a source of a certain kind. We can perceive sounds as having been produced by their sources in virtue of our experience (veridically) representing them as having been so produced.<sup>30</sup> As well as offering the best explanation of our experience of sounds and their sources, this description is consistent with the fact that our auditory system functions to extract information about the objects and events that produce the soundwaves it detects.

This account of sounds and their sources can be extended to include sequences of sounds. A sequence of grouped sounds is each experienced as having been produced by a source; they are experienced *as* grouped in virtue of being experienced as having been produced by the *same* source. We can explain, therefore, what it is to experience a sequence of sounds as grouped as an experience of them as having been produced by a single source. A sequence of sounds not experienced as grouped are experienced as having been produced by sources that are distinct.

This claim is plausible given my account of the function of auditory perception. Auditory perception functions to tell us about the sources of sounds; we experience sounds as grouped together because they are likely to have come from the same source and as the result of a process that extracts information about that source; that process produces experiences that represent both the source of the sounds and the sequence of sounds that it makes. My claim is that we hear the sequence of sounds as having been produced by the source we hear. When we hear the sounds made by footsteps, for example, we hear them as having been produced by a single object; when we hear an object as bouncing we hear a sequence of sounds produced by a

---

<sup>30</sup> Note that the explanation of what makes an experience of a sound veridical is different to the explanation of what makes an experience of the source of a sound veridical. That means that an experience may veridically represent a sound, but misrepresent the source of that sound. There are a number of auditory illusions that should be explained in just this way.

single object. In both cases, we hear the sequence as a grouped sequence because we hear them *as produced* by a single object.

I began with a description of auditory grouping as it occurs in a passage of music. Does the account I have given of auditory perception apply to our experience of grouping in music? We perceive music in virtue of the operation of the same mechanisms that enable us to perceive the sources of sounds.<sup>31</sup> The auditory system treats music as if it had been produced ecological events involving natural objects and groups frequency components accordingly. Our experience of grouping in music is the result of the auditory system making the best *ecological* sense of the frequency components it detects, and our experience of auditory grouping in music is the consequence. Composers of music have learnt to exploit perceptual mechanisms whose primary function is perception of the sources of sounds and to compose music that we experience as grouped in an aesthetically pleasing way.<sup>32</sup>

Auditory experience represents sounds as apparently produced by a source of a certain kind, that is, with certain properties; for the experience to be veridical the sound must have actually been produced by a source of that kind. This has the implication that our experience of sounds normally commits us to the existence of something other than sounds. That is surely right. Suppose that you hear the sound of a drum apparently being played in middle of the room. Your experience tells you that there is something happening there, that an event of a certain sort – the playing of a drum – is occurring. If there is no drum there, your experience has misled you. The experience wouldn't be veridical even if we contrived – using an array of speakers, for example – to reproduce exactly the sounds that a drum being played there would make. An experience produced in this way would be no more veridical than would be the visual experience of a perfect hologram of a vase on a table in front of you. A visual experience produced by a perfect hologram does not represent the world as it really is: it represents the existence of an object – a vase – that doesn't exist. Similarly, an auditory experience of a drum playing represents the existence an object; if there is no object being played there then your experience has misled you. It is because our auditory experience of sounds commits us to the existence of objects

---

<sup>31</sup> From the point of view of the function of the auditory system, musical sounds are produced in an abnormal way: naturally occurring objects and events – the kind of objects and events that the auditory system evolved to perceive – are very unlikely ever to produce the patterns of frequency components produced by groups of musical instruments playing in harmony.

<sup>32</sup> A point made by Deutsch 1999.

other than sounds that surround-sound systems in the cinema are so effective. Such systems use sounds to create the illusion of objects moving or being located around the listener. When you hear such sounds it seems as if objects really are moving past and around you. Knowing that the experiences are not veridical does not alter the effect: knowing that there are in reality no objects flying past does not prevent it seeming as if there are objects flying past. That it seems that there are objects flying past when we know that there aren't indicates that the illusion is perceptual and not the result of a judgment made on the basis of the experience.<sup>33</sup>

The claim that auditory experience represents sounds as having been produced by their sources can seem puzzling if we think that perceptual experience is restricted in what properties it can represent to those properties that determine how things perceptually appear.<sup>34</sup> Since having the property of being produced by a source of a certain kind is not a matter of a sound's having a certain appearance, how does our experience represent it as having that property? And since nothing other than sounds can auditorily appear to us, how can our auditory experience represent anything other than sounds? In particular, how can it represent the objects that are the sources of sounds?

If we think that perceptual experience is restricted to representing those properties that determine how things perceptually appear then our visual experience of objects can seem similarly puzzling. We see solid objects as solid objects and not just as the facing surfaces of solid objects, but how can visual experience represent something as actually being, say, cubic – as something with a rear surface – rather than merely having the *appearance* of being cubic – as a surface with the same appearance as that of a cube?<sup>35</sup> In representing something as cubic our visual experience represents it as having properties that go beyond the properties that actually determine how it appears.

---

<sup>33</sup> The illusion shows the immunity to judgment that is characteristic of the content of an experience as opposed to the content of judgement.

<sup>34</sup> There are two conceptions of appearance that are relevant here. Something can appear F if, taking our experience at face value, we would judge that it is F or something can appear F if it has the sensory quality of F-ness. Sometimes talk of appearance is shorthand for how someone would judge something to be; sometimes it stands for 'sensory' appearance. In the following discussion I mean it in the second sense.

<sup>35</sup> Since a solid cube can be visually indistinguishable from the facing surface of a cube – a cube from which every part not visible from the subject's point of view have been removed – having a rear surface and not being hollow are not properties that contribute to the appearance of a solid cube

Peacocke (1993, p.169) has claimed, surely correctly, that we experience objects as specifically material objects: a visual experience of a boulder in front of you produced by a perfect hologram of a boulder does not represent the world as it actually is, even if the hologram is visually indistinguishable from a real boulder. The content of the experience goes beyond the representation of the boulder's appearance – it represents the boulder as a material object; that is, as having the properties and causal powers that are essential to something's being a material object. Peacocke suggests that we can explain how someone can have a perceptual representation of a material object by supposing that their experience serves as input to a (perhaps only implicitly known) theory – an intuitive mechanics – whose theorems give content to their concept of a material object. Whether or not we accept the details of Peacocke's account, he is certainly right about two things. First, that visual experience represents objects as having properties that are not properties that determine how the object visually appears; and second, that an explanation of how visual experience can have such content will appeal to more general capacities of the subject – such as an intuitive understanding of mechanics – that are not perceptual capacities. What is true of the content of visual experience is also true of the content of auditory experience, and so whatever explanation we give of how visual experience can have content that represents material objects will also apply to auditory experience. The claim that auditory experience represents sounds as having been produced by their sources is, therefore, no more puzzling or problematic – and so no more objectionable – than the claim that visual experience represents objects as material objects.

Bregman, Albert S. 1990. *Auditory scene analysis: the perceptual organization of sound*. Cambridge, Ma.: MIT Press.

Bregman, Albert S, and J. Campbell. 1971. Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology* 89: 244-249.

Carello, C., J. B. Wagman, and M. T. Turvey. 2005. Acoustic specification of object properties. In *Moving image theory: Ecological Considerations*, edited by J. D. Anderson and B. Fisher. Carbondale, IL: Southern Illinois University Press.

- Deutsch, D. 1999. Grouping mechanisms in music. In *The psychology of music*, edited by D. Deutsch: Academic Press.
- Freed, D. J. 1990. Auditory correlates of perceived mallet hardness for a set of recorded percussive events. *Journal of the Acoustical Society of America* 87:311-322.
- Gaver, W. W. 1993. How do we hear in the world? Explorations in ecological acoustics. *Ecological Psychology* 5:285-313.
- Kunkler-Peck, A., and M. T. Turvey. 2000. Hearing shape. *Journal of Experimental Psychology: Human Perception and Performance* 1:279-294.
- Matthen, Mohan. 2005. *Seeing, Doing, and Knowing*. Oxford: Oxford University Press.
- McAdams, Stephen. 1993. Recognition of sound sources and events. In *Thinking in Sound*, edited by S. McAdams and E. Bigand. Oxford: Oxford University Press.
- Nakayama, Ken, Zijiang J He, and Shinsuke Shimojo. 1995. Visual Surface Representation: A Critical Link between Lower-level and Higher-level Vision. In *Visual Cognition Volume 2*, edited by S. M. Kosslyn and D. N. Osherson. Cambridge, Mass.: MIT Press.
- Neuhoff, John. 2004. Auditory motion and localisation. In *Ecological Acoustics*, edited by J. Neuhoff. London: Academic Press.
- Peacocke, C. 1982. *Sense and Content*. Oxford: Clarendon Press.
- . 1993. Intuitive mechanics, psychological reality and the idea of a material object. In *Spatial representation: problems in philosophy and psychology*, edited by N. Eilan, R. A. McCarthy and B. Brewer. Oxford: Blackwell Publishers.
- Peretz, Isabelle. 1993. Auditory agnosia: a functional analysis. In *Thinking in Sound*, edited by S. McAdams and E. Bigand. Oxford: Oxford University Press.
- Russell, M., and M. T. Turvey. 1999. Auditory perception of unimpeded passage. *Ecological Psychology* 11:175-188.
- Schiff, W., and R. Oldak. 1990. Accuracy of judging time to arrival: Effects of modality, trajectory, and gender. *Journal of Experimental Psychology: Human Perception and Performance* 16:303-316.
- Scruton, Roger. 1997. *The Aesthetics of Music*. Oxford: Oxford University Press.

- VenDerveer, N. J. 1979. Ecological acoustics: Human perception of environmental sounds, PhD thesis, 1979. Dissertation Abstracts International, 40, 4543B. (University Microfilms No. 80-04-002).
- Warren, Richard M. 1999. *Auditory perception: a new analysis and synthesis*. Cambridge: Cambridge University Press.
- Warren, W. H., and R. R. Verbrugge. 1984. Auditory perception of breaking and bouncing events: A case study in ecological acoustics. *Journal of Experimental Psychology: Human Perception and Performance* 10:704-712.
- Wildes, R., and W. Richards. 1988. Recovering material properties from sound. In *Natural computation*, edited by W. Richards. Cambridge, MA: MIT Press.