Auditory Perception and Sounds

It is a commonly held view that auditory perception functions to tell us about sounds and their properties. In this paper I argue that this common view is mistaken and that auditory perception functions to tell us about the objects that are the sources of sounds. In doing so, I provide a general theory of auditory perception and use it to give an account of the content of auditory experience and of the nature of sounds.

1. Introduction

The common view of auditory perception marks a distinction between it and visual perception: whereas the function of vision is to decipher visible cues to enable us to see objects in the world, it is supposed that the function of hearing is to decipher acoustic cues to enable us to hear sounds, which are those objects of experience that can be characterised in terms of their sensory qualities of pitch, loudness, and timbre. Whereas research that studies vision emphasises the recovery of distal properties of objects – properties such as movement, shape, and size – research that studies audition emphasises the recovery of the sensory qualities of sounds.[1]

According to this common view, then, the objects of auditory perception are sounds, and auditory perception functions to tell us about sounds and their properties. To the extent that we can perceive anything else about the world on the basis of hearing, it is because of a regular connection between sounds of certain kinds and the things that produce them. We can hear ducks, for example, in virtue of hearing quacking sounds and knowing, explicitly or otherwise, of the connection between sounds of that kind and ducks. It follows that an account of auditory perception

[1] In what follows I discuss only non-speech perception and the account I defend is intended to be an account of non-speech auditory perception. The problems of speech perception are such that it may best be treated as a distinct sense modality.

should be an account of how we perceive sounds and their qualities, and such has been the approach of those who have written on the topic.[2]

In this paper I reject this common view and argue that the function of auditory perception is, just like that of visual perception, to tell us about objects in our environment; to tell us, that is, about the objects and events around us that produce sounds rather than about the sounds they produce. I develop the argument for this by considering what account we should give of a relatively well-known auditory illusion.

In what follows I make the assumption that auditory experience, and perceptual experience generally, is intentional and can be characterised in terms of its representational content. Nothing of substance will depend on this assumption, and the view I describe is consistent with other accounts of perceptual experience. The common view of auditory perception is, in these terms, the view that auditory experience represents sounds and their properties; the view I shall argue for is the view that auditory experience represents the sources of the sounds it represents.

2. The pipe-organ illusion

One of the better known auditory illusions is the pipe-organ illusion. In the 18th century, pipe-organ manufacturers and players discovered that they could produce deep bass notes without the expense or space required for long bass pipes: if two notes of a fifth interval are played simultaneously, listeners hear a single note with a pitch an octave below that of the lower note of the dyad. For example, pipes measuring four feet and two feet eight inches can together produce a sound like that produced by a single eight foot pipe. The technique has been in use ever since.[3]

Hearing a sound produced in this way is not a case of hearing two sounds simultaneously; the pitch that one hears as a result of the combination appears to be the pitch of a single sound: one seems to hear a single sound with that pitch. This is not because we can never hear two (or more) sounds simultaneously. In general, we can hear two different, harmonically unrelated, pitches as distinct even when we

---

[2] Of whom there are few. See Casati and Dokic (1994) and O'Callaghan (2005); also O'Shaughnessy (1957a/b) and Pasnau (1999).

[3] For a more detailed history see organ historian Stephen Bicknell's website http://www.users.dircon.co.uk/~oneskull/3.6.01.htm.

cannot distinguish them by any other feature (such as their appearing to come from different locations or having different timbres).

Although there is *something* misleading or illusory about the experience produced by the pipe-organ illusion, it is not clear what. The illusion seems to show that it is possible for a single sound to be produced by two distinct and independent sources: two distinct sources produce a sound with a pitch distinct from that of the pitch of a sound that would be produced by either source alone. If that's right then two recent accounts of the nature of sounds must be mistaken.[4] But is that right? Does the illusion really produce the experience of a sound or does it involve an experience of a merely apparent sound, a sound that is not really there? Or does it perhaps involve an experience of a real sound as having a pitch it doesn't really have? Or is it illusory in some other way? It is, I think, far from clear both what we should say and why.

To answer these questions we need to know something more about why the illusion occurs. In general, perceptual illusions occur as a result of the way the perceptual system functions (and the way it can malfunction). Understanding why an illusion occurs sheds light on that function.[5] We can look to the function of a perceptual system in order to answer questions about what the experiences produced by that system represent, and so answer questions about the veridicality of particular experiences produced by it. Understanding why the pipe-organ illusion occurs will tell us about the function of the auditory system. If we know what the auditory system functions to represent then we will have grounds for deciding whether, in producing the illusion, it functions correctly and so produces a veridical experience, or incorrectly and so produces a non-veridical experience. Or so, anyway, I shall argue.

In the first half of this paper I give a characterisation, based on a variety of empirical data, of the function of auditory perception; in the second half I draw out the consequences of this characterisation. The arguments in the second half are plausible only in the light of the characterisation of the function given in the first half.

---

[4] For example the views of Casati and Dokic (1994) and Pasnau (1999).

[5] An example that illustrates this is the investigation of visual surface representation in Nakayama et al (1995).

This way of proceeding is necessary because it is not possible to understand the nature of auditory perception or of sounds independently of understanding the function of auditory perception, and the function of auditory perception is not well understood.[6] Although the empirical data to which I appeal are relatively uncontroversial, my interpretation is not; my aim, therefore, has been to provide sufficient empirical detail for the reader to be in a position to assess the plausibility of my characterisation of the function of auditory perception. A consequence of this is that the sections that follow are rather empirical; I ask the reader to bear with me until the second half when matters get more philosophical.

3. The information-bearing nature of soundwaves

The common view of auditory perception tends to assume that the sounds we hear can be physically characterised as soundwaves having a certain frequency and amplitude, and that hearing sounds involves simply detecting the frequency and amplitude of these soundwaves. This is almost entirely mistaken: it over-simplifies the nature of soundwaves and ignores the fact that they carry information about what produced them; consequently it both underestimates what perceptual processing the auditory system must do in order for us to hear sounds, and ignores the possibility that we might directly hear the sources of sounds and their properties.

To understand why this assumption is mistaken, we need do little more than consider the way a string vibrates. When a taut string is plucked it vibrates along its entire length with the maximum displacement occurring in the middle of the string.[7] The wavelength of this vibration is twice the length of the string; its frequency[8] – the

---

[6] Unlike the function of visual perception, there is little agreement about what auditory perception functions to do – what it is for. This is, I suspect, because research has focused on speech and music perception which, although important, are not what auditory perception is for (cats have an acute sense of hearing, but perceive neither speech nor music).
[7] Plucked strings behave slightly differently to strings caused to vibrate by other means. For a detailed discussion see Fletcher and Rossing 1998, secs. 2.7 – 2.11.
[8] The frequency with which the string vibrates is inversely proportional to the wavelength, and proportional to the velocity of the wave. The velocity of the wave depends on the tension of the string and its mass; changing the tension therefore changes the fundamental frequency of the vibration and the pitch of the sound produced.

lowest frequency component of the string's complex vibration – is known as the fundamental frequency of the vibration. The string also vibrates at three times the fundamental frequency, with a wavelength which corresponds to one and a half times the length of the string – imagine the string divided in three with each part vibrating – at five times the fundamental frequency – divide the string into five – and so on. These higher frequencies – the odd integer multiples of the fundamental frequency – are known as *harmonics* or *partials* of that fundamental. The overall vibration of the string is both complex – the result of adding together or superimposing these different component vibrations[9] – and changes over time as energy is lost and the vibration decays. Different frequency components may decay at different rates so that the spectral composition of the vibration changes over time.

Most of the sounds we hear are produced by events involving the interaction of objects rather than by vibrating strings. We hear sounds produced by things tapping, knocking, banging, rubbing and scraping against one another. In much the same way that a string vibrates when it is plucked, an object, when struck, vibrates in a complex way that comprises many different frequency components.

When an object is stuck the force of the impact deforms it; once the force is removed, the object vibrates until the energy of the deformation is lost and the object returns to its equilibrium state. The character of an object's vibration – the number and proportion of different frequency components and the way they change over time – is determined by its physical properties. The size of an object will determine the lowest frequency at which it vibrates. A solid object, unlike a string, vibrates along several dimensions, and its shape and size will determine both the frequency and spectral composition of its vibration. The time it takes for the object to return to equilibrium is determined by how the vibration is damped – that is, how quickly it loses energy. Heavily damped vibrations decay rapidly, whereas lightly damped

[9] Two sine waves can be combined to produce a complex waveform which is simply the result of summing the amplitudes of the two waves at each moment of time. Complex waveforms can be combined in the same way. Conversely, *any* complex waveform can be analysed into a number of component sine waves of various frequencies and amplitudes which, when added together in the correct phase, produce the analysed wave (this process of analysis is known as Fourier analysis). The complex sound wave that is detected by our ears is equivalent to a set of phase related sine wave components of differing frequency and amplitude. Many psychoacoustic theories suppose that the auditory system must perform some equivalent of a Fourier analysis.

vibrations are prolonged. Damping may vary with frequency so that different frequency components decay at different rates; the kind and amount of damping therefore affects the way the vibration varies over time; in particular, it affects the spectral composition of the vibration over time and not just its average amplitude.[10] All these features of an object contribute to the pattern of frequency components of the vibration and to the way that pattern changes over time. The kind of interaction between objects that causes the vibration also affects the character of that vibration. Whether the object was struck once or continuously, whether it was scraped, and so on, affects the time-course of the subsequent vibration. The force with which an object is struck affects the spectral composition as well as the amplitude of the resulting vibration; typically, the relative intensity of the higher frequency harmonics of a vibration increases when an object is struck with greater force.[11] The character of the vibration produced by an object's interactions with other objects is, therefore, partly determined by its physical attributes and by the nature of its interactions. Because of this, the vibration embodies or carries information about the object's physical attributes; information concerning, for example, the size or mass of the object, the kind of material of which it is composed, its density; and it carries information about the object's interaction with other objects, about the force with which it was struck, the number of times it was struck, and whether the blow was clean.[12]

When the object is immersed in a suitable medium, its vibration produces a compression wave in that medium. The compression wave produced by a vibrating object interacts with other objects in the environment, is filtered by passing through the medium, and is differentially reflected by surrounding surfaces; this alters the spectral composition of the wave in determinate ways. The compression waves that

---

[10] How much damping different materials produce very clearly affects the character of the sound an object makes when it is struck – for example, wood, which is heavily damped, makes a thunking sound, whereas metal, which is less damped, rings.

[11] See Chowning 1999, p.270.

[12] Although for the sake of simplicity I have described how differences in geometrical properties of objects affect the way that they vibrate, they do so only in virtue of being correlated with the mechanical properties of those objects; it is likely that the auditory system detects or tracks mechanical rather than geometrical properties of objects. Further work needs to be done to discover which mechanical properties the auditory system detects or tracks. See Carello et al (2005, pp.14 ff).

reach the ears therefore carry information about the physical space or environment within which the events that produced the wave occur.

In virtue of having been produced by objects, and structured by interactions with their surroundings, the soundwaves that reach us embody a great deal of information about object-involving events occurring in our environment, about the number and properties of these objects, and about the environment in which these events occur. In picturing auditory perception as the perception of sounds, the fact that soundwaves embody this information is either overlooked or ignored. The information is, at least in principle, detectable, and although little systematic research has been done, it is evident that our auditory system can detect it: it is evident, that is, that we can hear the sources of sounds and their properties.

4. We can perceive sources

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 – it is more likely, though, that you will attend to what is making the sound: to the fact 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 sound.[13] 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

[13] When they do this, listeners adopt what Gaver (1993a) calls a 'musical' and Scruton (1987, pp.2 ff.) an 'acousmatic' attitude to what they hear.

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.

There is evidence that when we attend in this way we can perceive sound producing events and objects and their properties, and are capable of recognising very specific characteristics of the events and objects we hear.[14] 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.[15] 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.[16] 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)'".[17]

---

[14] 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).

[15] Listeners' success rate was 99%; see Warren and Verbrugge, 1984.

[16] The success rate was about 95%; see VanDerveer, 1979.

[17] 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

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,[18] and the time to contact – that is, the time at which we will collide – with a sound source that is moving towards us.[19] We are good at hearing whether an invisible object making a noise is within reach;[20] 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.[21] We can identify the material composition of an object from the sound of an impact,[22] and perceive the force of the impact.[23] 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.[24] 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.[25]

What these examples suggest is that our auditory system is able to extract the information about objects embodied in sound waves. The resulting experience represents properties of the sources of sounds as well as sounds produced by those

---

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.

[18] Neuhoff, 2004.
[19] Schiff and Oldak, 1990.
[20] Carello et al., 1998.
[21] Russell and Turvey, 1999.
[22] Wildes and Richards, 1988.
[23] Freed, 1990.
[24] 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).
[25] Carello, et al., 1998.

sources. It is in virtue of our experience representing properties of the sources of sounds we are able to recognise and discriminate those sources in the way I have described. The examples suggest that our auditory experience can represent sounds as produced by things of a certain size, by something rolling, by an object striking a surface, or even by hands clapping; that we may experience a sound as made up of a number of simpler sounds and hear the sequences as having been produced by footsteps, or breaking glass; and, although these particular examples are silent on the issue, there are other examples which show that we can perceive aspects of the environment in which the sound was produced – that we can experience a sound as being produced by an object striking a surface in an enclosed space, for example.

So far I have argued that soundwaves carry information about the things that produced them and that, in virtue of this, we can perceive those things – that our auditory system produces experiences that represent the sources of sounds and their properties. In the following sections I will describe the connection between these facts and our experience of sounds.

5. The problem of source perception

If the sounds we heard were only ever produced by one event at a time, and if the transmission of soundwaves through a medium were more robust, 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. There may be, however, and often are, many different events producing sounds simultaneously. Compression waves produced by these different events interact with each other, with surrounding objects, and with the medium, so that the compression wave that reaches and 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 many different frequency components produced by different events. Not only are the frequency components detected by the ear at any moment the product of different sound producing events, those from a single event, as well as travelling directly, may reach the ears indirectly after having been reflected from other surrounding objects and surfaces, or being otherwise distorted. The fact that frequency components are detected simultaneously is therefore no indication that they were produced simultaneously by a single event. And the fact that frequency components are detected successively may indicate that they have been produced by temporally successive parts of a single event, by two distinct but consecutive events, or be the result of a component produced by a single event being detected twice having reached the ear directly and then later indirectly after having been reflected from a nearby object or surface. Furthermore, the frequency components produced by different events, or reflected components produced by a single event, may constructively or destructively interfere with, obscure or mask one another.

The auditory system can detect only the patterns of frequency components that make up the complex vibrations of the soundwaves that reach the ears. It must construct a representation of the environment by extracting the information that this pattern embodies. Auditory perception must, therefore, involve 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 of 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 auditory process in part (though only in part)[26] involves a grouping or organising of frequency components in such a way as to produce the experience we have of sounds. Grouping detected frequency components in such a way as to extract information about the environment is a far from trivial task. Its difficulty is illustrated

[26] The particular grouping processes that I describe in this paper are only a small part of the process or processes that produce our auditory experience. Such grouping explains why we hear the sounds we hear, but doesn't explain how information about the sources of sounds is extracted. Furthermore, the grouping processes that I discuss are only a subset of those that occur: other processes are responsible for grouping over time, grouping according to schemata, and so on.

by Albert Bregman's imaginary game: suppose that you are standing by a lake on which there are boats:

> The game is this. Your friend digs two narrow channels up from the side of the lake. Each is a few feet long and a few inches wide and they are spaced a few feet apart. Halfway up each one, your friend stretches a handkerchief and fastens it to the side of the channel. As waves reach the side of the lake they travel up the channels and cause the two handkerchiefs to go into motion. You are allowed to look only at the handkerchiefs and from their motions to answer a series of questions: How many boats are there on the lake and where are they? Which is the most powerful one? Which is the closer? Is the wind blowing? Has any large object been dropped suddenly into the lake?
>
> Solving this problem seems impossible, but it is a strict analogy to the problem faced by our auditory systems. The lake represents the lake of air that surrounds us. The two channels are our two ear canals, and the handkerchiefs are our ear drums. The only information that the auditory system has available to it, or ever will have, is the vibrations of these two ear drums. Yet it seems able to answer questions very like the ones that were asked by the side of the lake: How many people are talking? Which one is louder, or closer? Is there a machine humming in the background (Bregman 1990, pp. 5-6).

Answering these questions about events occurring in the environment requires that frequency components detected by the ears be grouped in such a way that those produced by a single source are treated together and distinguished from those produced by distinct sources. This grouping is necessary in order for our perceptual system to extract information about the sources of sounds and so produce auditory experiences that represent those sources. How is such grouping achieved?

6. How is grouping achieved?

Considered in isolation, a single frequency component carries very little information about its source. There is nothing intrinsic to a particular frequency component that marks it as having been produced by one event rather than another, and nothing intrinsic to each of a set of components that marks it as having been produced by a single event simultaneously with other components.[27] How then does the auditory system group frequency components?[28] Part of the explanation depends on the fact that there will be relationships between sets of components that have been produced by the same event that are unlikely to exist by chance. These relationships obtain in virtue of the physical properties of the different frequency components – properties such as the timing, frequency, and amplitude of the waveform – and in virtue of the differential effects different components have on the two ears.

I have described (in section 3) how an object's vibration involves frequency components that are harmonics of a fundamental frequency. A consequence of this is that the frequency components of a soundwave produced by a single object will be harmonically related. These harmonic relationships are unlikely to exist between frequency components produced by distinct objects, 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 they are likely to have been produced by the same source event. Similarly, the soundwave produced when an object is struck will have frequency components which share temporal properties: all the components will begin at the same moment in time. If the auditory system detects a number of frequency components which have the same temporal onset then it is likely that those components have all been produced by the same source event. Components produced by such an event are likely to be in phase with one another, those produced by distinct events unlikely to be so; if components are detected that are in phase with one another they are likely to have been produced by the same source event. The frequency of components produced by an event may change or modulate over time. The frequencies of all the components produced by a single event will tend to change in the same way, but it is very unlikely that components produced by distinct events will share a pattern of frequency modulation. Therefore, if a set of components is detected that have a common pattern of frequency modulation then they are likely to have been

[27] Compare this to the question: How does the auditory system know which *sequences* of frequency components have been produced by the same event? The answers to both questions are not entirely independent, but for the sake of brevity I am omitting any discussion of sequential integration.

[28] Relatively little is known about the details of how this integration is achieved. This is especially true for naturally produced sounds (glass breaking, water flowing, etc). In what follows I sketch some of the principles involved.

produced by the same source event. A soundwave produced by one object scraping against another will have frequency components whose amplitudes vary simultaneously over time in a way that those produced by distinct events will not; if the auditory system detects components with a common pattern of amplitude modulation then they are likely to have been produced by the same source event. And so on.

These are all examples of relationships that are only likely to exist between components produced by a single event and are unlikely to exist otherwise. In grouping or organising frequency components, the auditory system is able to detect and exploit these relationships: when frequency components are detected that bear these relationships to each other they tend to be grouped together and treated by the auditory system as having been produced a single source event.[29]

This brief list of relationships is not intended to be exhaustive; there are others that the auditory system can exploit. In particular there are relationships that exist between components over time which are involved in their sequential organisation – in grouping together a component at one time with a component at a later time – and there are relationships that can be exploited by top-down processes – by processes that draw on knowledge of the properties of object or event that produced the sound. Such top-down processes are likely to be involved in the perception of any temporally structured event, including speech perception, and in the perception of familiar 'meaningful' sounds, such as the sound of a dog's bark, of footsteps, of a car's engine, and so on.[30]

What all these examples of how frequency components are grouped show, both those that involve simultaneous grouping and those that involve sequential grouping, is the following: that we cannot explain why the auditory system groups frequency components in the way it does except in terms of a process whose function

---

[29] As I am using it, the term 'grouped' is the name of a process rather than the product of a process. That the auditory system tends to group components that share these features has been experimentally demonstrated. My discussion here draws on Bregman (1990), especially ch.3, to which the reader should refer for details of the empirical support for the claims in the text.

[30] The relationships I have described are all involved in what Bregman calls 'primitive stream segregation' (1990, pp.38 ff.); this is a process that is likely to be innate and which exploits invariable properties of the subject's environment. It contrasts with what he calls 'schema-based segregation' (1990, pp. 395 ff., and pp.665 ff.) which is learned.

is to group together all and only those components that have been produced by the same source event. Considered independently of their sources, the way the auditory system groups frequency components is arbitrary. It is only relative to their sources that grouping makes sense.

Why does the auditory system function to group together all and only those components that have been produced by the same source event? I have described how soundwaves, in virtue of having been caused and systematically structured by events occurring in our environment and by the environment itself, carry information about those source events and the environment. Extracting this information requires the auditory system to determine the number of sources that are producing the components it detects; this in turn requires the auditory system to group components according to whether they have been produced by the same source event. Since information about the event is carried in the pattern or structure of frequency components it produces, the auditory system must group just those components produced by that event in order to recover the information. Grouping is, therefore, a necessary step in a process that extracts information about the sources of sounds; it is a necessary part of a perceptual process which functions to produce experiences that represent the sources of sounds.

Although grouping components from the same source is necessary for the auditory system to perform the function of representing the sources of sounds, it's not otherwise necessary. We can imagine an auditory system – that is, a system that detects the frequency components of soundwaves – that functioned to represent, say, acoustic spectra (to produce a spectrograph of what it detects) or that functioned simply to group frequency components in an aesthetically pleasing way. Although it is difficult to imagine circumstances in which such perceptual systems would be useful to a creature, they are not conceptually incoherent.

7. Grouping and our experience of sounds

What is the connection between auditory grouping and the sounds we experience? In talking of grouping of frequency components by the auditory system I am describing a functional process; in talking of a set of components as having been grouped I mean that they are treated by subsequent auditory processing as belonging to a group. Our auditory experience is representational and sounds are the objects of representational states. What sounds our experience represents and how it represents them to be is determined by how the auditory system groups the frequency components it detects. In particular, the representational content of our experience is determined in such a way that we experience a sound corresponding to each grouping of frequency components. If all the frequency components detected by the auditory system are grouped together then we have an experience as of a single sound whose character is partly determined by the components that have been grouped in producing our experience of it; if those same components were grouped into two, we would have an experience as of two sounds. The evidence for this is empirical. By manipulating the properties of frequency components we can change the way they are grouped and so change the number and character of the sounds a listener experiences.[31]

For example, if played a pure tone – in effect a single frequency component – a listener will have an experience as of a single sound. If, after a short interval, a second tone at a different frequency is added to the first, a listener will typically hear one of two things. If the second tone is at a frequency unrelated to the first, then she will have an experience as of a distinct sound; that is, her experience will be as of two sounds. This happens because the auditory system has grouped the two frequency components separately. If the second tone is at a frequency related to the first by being, say, its first harmonic, then the subject is unlikely to experience it as a distinct sound; she may, rather, experience a slight change in the character of the initial sound. This happens because the auditory system treats both components as having been produced by a single source and so groups them together. The same thing happens if a third tone is added; and so on. As consequence of their being grouped together such a set of frequency components produces an experience as of a single sound. If, however, small random frequency variations are added to subsets of the components, a listener will typically have an experience of distinct and countable sounds. Such frequency variations are sufficient to prevent the components being grouped by the

[31] Many of the same principles apply to grouping in music, and Diana Deutsch (1999) describes several examples of how changes in grouping change what musical sounds, and what sequences of musical sounds, are heard. It is an interesting question whether our experience of the longer sequences of tones in a melody can be explained in terms of mechanisms that evolved for the perception of ecological (non-musical) events.

auditory system as having been produced by a single source.[32] A particularly striking example of this phenomenon is provided by the sound made by striking a bell.[33] Normally, listeners experience this as a single sound. When different random variations in frequency are added to three different sets of harmonics of the soundwave, the sound appears to split into three. When these random variations are removed, the three sounds appear to merge back into a single sound. In this example, artificially altering the soundwave changes the way the auditory system groups different frequency components and, as a consequence, changes the number and character of the sounds a listener experiences.

Given that the auditory system functions to group together all and only those frequency components that have been produced by the same event, and that what sounds our experience represents is determined by how the auditory system groups frequency components – so that a represented sound corresponds to a grouping of frequency components – it follows that the auditory system functions in such as way as to produce experiences that represent sounds which correspond to the events which produced them; which correspond, that is, to their sources. We cannot, therefore, explain why we experience the sounds we do except in terms of a process which functions to produce experiences of sounds that correspond to the events which produced them. That is, we cannot explain our experience of sounds except in terms of their sources.

This reverses the order of explanation implicit in the common view of auditory perception that I began by describing. According to the common view, it is because we perceive the sounds we do that we can come to know anything about the sources of sounds: we hear a sound and recognise it to have been produced by a certain kind of source. We explain our experience or perception of the sources of sounds in terms of our perception of sounds. But such an explanation would only be possible if were possible to explain why we experience the sounds we do without reference to a process that functions to enable us to perceive their sources. No such explanation is possible. We perceive sounds because we perceive the sources of sounds.

---

[32] Chowning, 1999, pp. 265 ff.
[33] An example recording is available at ****.

I began by characterising the common view of auditory perception according to which auditory perception functions to tell us about sounds and their properties; to the extent we can perceive anything else about the world on the basis of hearing it is because of a regular connection between sounds of certain kinds and the things that produce them.

In the first half of this paper I have drawn on a range of empirical data to argue that the function of auditory perception is to tell us about objects in our environment, rather than about sounds. This conclusion follows from the fact that we can only explain why we experience the sounds we do in terms of the part they play in the process which carries out the function of telling us about objects. Our experience of sounds, rather than being the goal of that function, is the result of the operation of a process that implements the function. The sounds we experience are determined by an intermediate stage in this process. Although the auditory system functions to tell us about sounds in the sense that it tells us about sounds as part of its functioning, telling us about sounds is not its function. Telling us about sounds is not the goal of audition: its goal is to tell us about objects.

This conclusion is important and I have drawn on three kinds of empirical evidence to reach it. Firstly, evidence about the connection between the physical and other properties of objects and the way they vibrate; secondly, evidence that the auditory system can extract information about the sources of sounds from vibrations transmitted by soundwaves; thirdly, evidence about the operation of the process that extracts this information.[34] Although based on empirical evidence, my characterisation of the function of auditory perception is an interpretation of that evidence;[35] and it has various philosophical consequences, not least for what we should say about the content of auditory experience and about the nature of sounds. I

[34] The significance of this evidence for accounts of auditory perception is not always appreciated by those doing empirical work in audition. They tend to study one of these areas in isolation from the others. But surely it would be fruitful for those working on the perception of music, for example, to relate the properties of sounds that are said to determine grouping in music to the processes for object perception that I have been sketching. It is more plausible to suppose that perception of music emerges from a general capacity for auditory object perception (of the kind I have described) than to think we have evolved an independent capacity for the perception of music.

[35] My interpretation is, as far as I know, novel.

explore some of these consequences in the second half of this paper. Before doing that, however, I want to say something about errors and the pipe-organ illusion.

8. Grouping errors

The existence of various relationships between the frequency components detected by the ears, some of which I described in section 6, makes it probable that those components have been produced by the same source event. We can, therefore, view these relationships as constituting evidence that the components have been produced by a single event, evidence that the auditory system uses in determining how to group components together and whether to treat them as having been produced by a single source.

In ideal circumstances this evidence will be unequivocal and will indicate a single way of grouping detected components. In less than ideal circumstances, the evidence may be more equivocal.[36] Soundwaves suffer interference during transmission, and individual frequency components may become obliterated or distorted; in noisy environments some components may be masked by others; damage or deterioration to the ears may mean that some components are not detected. In such circumstances the evidence may not mandate a single way of grouping detected frequency components; some evidence may favour one way of grouping and other evidence a different way of grouping. When this happens, the auditory system may disregard some evidence to make the best sense it can of components it detects; we can think of the sources represented by the resultant experience as providing the best – or most likely – explanation of the pattern of detected frequency components. Making best sense of the evidence will normally result in a correct way of grouping,

[36] With the advent of recorded and electronically produced sounds, non-ideal circumstances (from the point of view of the function of the auditory system) have become common. Soundwaves and frequency components that are never likely to occur naturally and that will not have occurred during the evolutionary history of the auditory system are now commonplace. Sounds played over stereo loudspeakers are a good example: soundwaves produced by two sources have relationships that are practically impossible in nature and as a consequence, although they are produced by two sources, the auditory system ignores or disregards their spatial discrepancies and treats them as having a single (and merely apparent) source.

producing an experience both of sounds that corresponds to their sources, and that veridically represents those sources.[37] Sometimes, however, the grouping is incorrect and results in an experience of sounds that do not correspond to events which produced them.

In non-ideal circumstances, then, the auditory system may group frequency components incorrectly; that is, components produced by different sources may be grouped together; those produced by a single source distinguished into distinct groups; or components from one source grouped with those from another. Such groupings are incorrect relative to the auditory system's function of grouping all and only frequency components produced by the same source. This incorrect grouping produces an experience of sounds that do not correspond to events which produced them.

The experience produced by the pipe-organ illusion is the result of this kind of incorrect grouping. The auditory system has evidence that suggests that the frequency components produced by the two pipes have a single source: air in the two organ pipes vibrates at frequencies that are harmonically related producing a soundwave with harmonically related frequency components; the two pipes are played simultaneously to produce frequency components with synchronous onsets; and the pipes are made from the same material so produce frequency components which are likely to be micro-modulated in similar ways. It is very unlikely that distinct naturally occurring events would produce soundwaves with frequency components that are related in this way. The auditory system therefore treats these frequency components – that have in fact been produced by two distinct events – as having been produced by a single event, disregarding any evidence – concerning, say, the locations of the sources – that may conflict with that interpretation. This produces an experience that represents a sound as having been produced by single source event, which is the best – most likely – explanation of the frequency components detected.

That, however, cannot be the whole story. The incorrect grouping of frequency components explains why the pipe-organ illusion produces an experience of a single sound; it doesn't explain why that sound is experienced as having a pitch

[37] Often in the evolutionary history of the auditory; it may no longer be true.

which is different to the pitch of the sound that would be produced by either pipe playing alone. That can only be explained by appealling both to how the auditory system makes best sense of the evidence it detects, together with an account of pitch perception.

The auditory system tends to group harmonically related frequency components because they are unlikely to exist other than as having been produced by a single event. Suppose, however, that only a partial set of harmonics is detected – a set of higher harmonics, for instance, but no fundamental frequency. There are two possible explanations of the existence of such a set. The first is that two or more distinct sources are vibrating at frequencies with fundamentals identical to the first, second, and higher detected harmonics respectively. Such circumstances are very unlikely to occur naturally though it is, of course, exactly what happens to produce the pipe-organ illusion. The second, and far more likely, explanation of the detection of just the higher harmonics is that the fundamental component produced by single source was not detected because it was masked, had been filtered out of the soundwave, or otherwise obscured. Because this second explanation is more likely, the auditory system treats a partial set of harmonics in the same way as the same harmonics would be treated were the fundamental to be simultaneously detected. This is true even if only some of the higher harmonics are detected.

Again, this makes sense given that the auditory system functions in the way that I have described. The auditory system groups components as part of a process whose function is to extract information about the source of the sound, and so produce experiences that represent the source. Harmonics embody information about the number of sources – two or more different sources would be required to produce sets of harmonics with different fundamental frequencies – and – because the frequency of the fundamental of a set of harmonics is determined by the physical properties of the object, in particular its size – information about properties of the source. By in effect ignoring the fact that the fundamental is missing, the auditory system both groups the detected components in a way that normally corresponds a source, and recovers the information about the properties of the source which the harmonic structure of that grouping embodies. Grouping in this way will, in normal circumstances, produce a veridical experience of the source of the sound.

How is this connected to the pitch a sound is experienced as having? Pitch is the auditory feature of sounds in virtue of which they can be ordered on a scale from high to low. Many, though not all, sounds are experienced as having a pitch. Sounds produced by simple vibrations, producing soundwaves with a single frequency component, are experienced as having a pitch determined by the frequency of that component. Most of the sounds we hear are produced by complex vibrations made up of many different frequency components. We nonetheless experience many such sounds as having a pitch. For example, we normally experience the sound produced by a plucked string as having a particular pitch. Since there is no single frequency which is *the* frequency of the string's vibration, the pitch we hear the sound to have is not simply determined by, or identical with, the frequency of the vibration that caused our experience of it.

When the auditory system groups a set of simultaneously detected harmonics to produce an experience of a single sound, the pitch such a sound is experienced to have is determined by the fundamental frequency of the grouped harmonic components.[38] For example, a soundwave with frequency components at 200Hz, 400Hz, 600Hz, and 800Hz normally produces an experience of a sound having the same pitch as a sound produced by a soundwave with a single 200Hz component. Since, even when the fundamental frequency of a set of components is absent, the auditory system will tend to group the harmonics and treat them in the same way as a set with the fundamental present, a set of harmonics with a missing fundamental produce an experience of a sound with a pitch the same as that of the sound produced when the fundamental frequency is present. For example, a soundwave with frequency components at 400Hz, 600Hz, and 800Hz normally produces an experience of a sound having the same pitch as the experience of a sound produced by a soundwave with a single 200Hz component. This explains why when we musical instruments over the telephone or on a small radio – neither of which is able to reproduce low frequencies – we hear them as having their normal pitch.[39]

---

[38] Bregman calls this the 'harmonicity principle' (1990, pp. 232 ff).

[39] Whilst it has been long known that the perceived pitch of a sound is determined by its harmonic structure, in particular by its (perhaps missing) fundamental frequency, there is still no widely accepted explanation of how the auditory system does this. Most textbooks on psychoacoustics contain a description of what are taken to be the most plausible theories (but see the next footnote). For a summary and further reading see, for example Gelfand (1998, ch.12). For an alternative temporal model of pitch perception see Griffiths et al. (1998).

The explanation of our experience of the pitch of the sound in the pipe-organ illusion thus involves two steps. The first appeals to the fact that the auditory system groups frequency components that are harmonics of a common fundamental; the second to the fact that the auditory system treats a set of higher harmonics of a missing fundamental in the same way it treats a set of harmonics when the fundamental is present. The two organ pipes produce frequency components which are harmonically related – such as would normally be higher harmonics of a missing fundamental – and these are treated by the auditory system in the same way it would a set in which the missing fundamental was present – a set that could only be produced by a much larger pipe – so as to produce an experience of a sound with a pitch the same as that of a sound produced by a much larger pipe. Both parts of this explanation appeal to a process which functions to produce experiences which represent the source of the sound; a process that also produces an experience of a sound with the same pitch as that which would be produced by a source with different properties.[40]

Because the best explanation of the detection of a partial set of harmonics is that they were all produced by a single source, treating a partial set of harmonics as harmonics of a single source will normally produce an experience that veridically represents the source. Therefore, in non-ideal circumstances when only a partial set of harmonics is detected, the auditory system normally functions to veridically represent the source of those harmonics. In the case of the pipe-organ illusion, the

[40] There's an assumption in the psychoacoustics literature that the pitch of a sound is identical to its fundamental frequency; that we, in some sense, *experience* this frequency or that this frequency produces a *sensation* of pitch. That makes it puzzling why – indeed *how* – a set of harmonics with a missing fundamental is experienced as having the same pitch as a set with the fundamental present: in the absence of the frequency component, how could we have the experience? But the assumption (and the corresponding puzzle) is just mistaken. The auditory system is representational; it represents sounds as being some way, and the sources of sounds as being some way. The pitch a sound is experienced as having is determined by how the experience of the sound represents that sound to be; that is, by what pitch it represents it as having. What pitch a sound is represented as having may be determined by relatively complex properties of the auditory stimulus, and be the result of cognitive processing. We needn't think that in experiencing a sound we are directly aware of frequency components. Similarly, there is little reason to think that, in representing a sound as having a particular pitch, our auditory experience is simply representing the presence of a particular frequency component rather than, say, a pattern amongst components that can be instantiated even in the absence of a particular component. A nice example is provided by pitch changes due to the Doppler effect which are usually explained by appealing to changes in the frequency of a sound wave. Whatever you might think, the pitch shift cannot be explained that way – it is a cognitive rather than a sensory effect (McBeath and Neuhoff 2002).

auditory system produces an experience of the source of what would normally produce the frequency components it detects. Because these frequency components were produced in an abnormal way, the experience produced is not veridical: rather than representing what actually produced them, it represents what would normally produce those components, namely a single and larger source.

9. What are sounds?

In the preceding sections I have described how the auditory system groups frequency components to produce experiences as of sounds corresponding to their sources, and how this grouping process may go wrong and produce experiences as of sounds that do not correspond to their sources. In both cases, the auditory system produces experiences that represent sounds. When are these experiences veridical? In order to answer that question we need to know what, in representing a sound, our auditory experience is representing – we need to know what the correctness conditions are for experiences of sounds.

When our auditory system functions normally – that is, when it functions to produce experiences that veridically represent the *sources* of the sounds it represents – our experience of sounds depends counterfactually – in the way that I have already described – on patterns or structures of frequency components instantiated by the soundwave that reaches and is detected by the ears. It's in virtue of the frequency components of a soundwave instantiating a certain pattern or structure that they can be grouped together by the auditory system to produce an experience as of a sound. Given this dependence, the following claim is plausible: in representing a sound our auditory experience is representing an instance of a pattern or structure – the pattern or structure of frequency components that would normally produce an experience of that sound. It follows that an experience as of a sound is veridical only if it is produced by the instantiation of pattern or structure of frequency components that would normally produce that experience; it is not veridical if either it is not produced by any such pattern or if it is produced by a pattern that would not *normally* produce that experience.

Hallucinatory experiences of sounds are non-veridical in the first way. A hallucinatory experience represents a sound – a pattern instantiated by frequency components – that does not exist – the experience is not produced by an instantiation of that pattern or indeed any pattern. Experiences that are non-veridical in the second way are most likely to occur as a result of damage to the auditory system. Lesions within the central auditory system can affect our experience of sounds in such a way that sounds are experienced as altered in volume, in tone or timbre.[41] In such cases, as a result of cortical damage, the experience of a sound is produced by the instantiation of a pattern of frequency components that would not normally produce that experience.

An experience of a sound can be produced by an instantiated pattern of frequency components that would normally produce that experience even when the instance of the pattern itself is not produced in the normal way (from the point of view of the function of the auditory system). This happens with most artificially produced sounds. Stereo loudspeakers, for example, playing a recording of several objects being dropped onto a hard surface, produce a soundwave that instantiates a pattern of frequency components that would normally be produced by several objects striking a hard surface. Loudspeakers, from the point of view of the function of the auditory system, are an abnormal way of producing such a soundwave. This abnormally produced soundwave instantiates a pattern that nonetheless produces an experience as of sounds normally: the experience as of sounds is the same as that such an instantiated pattern would normally produce. It is an experience that represents the instantiation of a pattern or structure of frequency components that would normally produce an experience of those sounds; it has been produced by such a pattern and so is veridical. Although the experience veridically represents the sounds, because the instance of the pattern that produces the experience was not produced normally it produces an experience of sounds that do not correspond to their sources. The resultant experience therefore misrepresents the sources of the sounds: it represents the sounds as having been produced by sources that did not produce them.

This is exactly what happens in the pipe-organ illusion. A soundwave that instantiates a pattern of frequency components that would normally be produced by a single source is in fact produced by two sources – two different organ pipes. This results in an experience of a single sound that seems to have been produced by a

[41] For a survey of the effects of brain lesions on auditory perception see Griffiths et al. (1999) and Griffiths (2002a, b).

single source. To someone who hears the sound there will seem to be a single event or object producing it, when in fact there are two. Although the experience – in virtue of being produced by an instantiated pattern of frequency components that would normally it – veridically represents the sound, it misrepresents that sound as having been produced by a source that didn't produce it. The experience misrepresents the source of the sound.

I suggested that we should think of our experience of sounds as representing the patterns instantiated by frequency components, on which they counterfactually depend. The frequency components that instantiate such patterns are the proximal cause of our auditory experience. It might, therefore, be objected that although in general a perceptual representation counterfactually depends on its proximal cause, it doesn't follow that it represents that proximal cause. Our visual experience, for example, counterfactually depends on its proximal cause – patterns of light detected by the retina – but it would be a mistake to think it represents those patterns. In claiming that our auditory experience as of a sound represents an instance of a pattern or structure of frequency components aren't I mistakenly assuming that an auditory experience represents its proximal cause? In fact, since the medium through which soundwaves travel simply transmits vibrations, it might seem reasonable to suppose that when our auditory system functions normally and our experience of the source of the sound is veridical, whatever patterns of frequency components are instantiated by the soundwave that reaches the ear will be (or will have been) instantiated by the objects whose vibration produced the soundwave. In these cases our experience of sounds depends counterfactually on frequency components instantiated by the source of the sound, so isn't it more plausible to claim that our experience of sounds represents patterns instantiated by the sources of sounds, rather than by soundwaves? It would follow that sounds are instantiated by the objects that produce them. In that case our experience of sounds will be veridical just in case our experience of the source of the sound is, and if our experience misrepresents the source of a sound it must misrepresent the sound.

According to the objection, the fact that our experience of sounds depends on the pattern of frequency components instantiated by the soundwave just tells us what *causes* our *experiences* of sounds, it doesn't tell us what those experiences represent and so it doesn't tell us when the experiences are veridical. So what justifies the move from the claim about what determines our *experiences* of sounds to the claim about what *sounds* we experience?

The auditory system function that I have characterised describes the relationship between three things: objects and events involving objects; the objects' vibrations (which can be characterised as a set of frequency components instantiated by the object); and the vibrations of the soundwave (which can be characterised as a set of frequency components instantiated by the soundwave). According to my characterisation, the auditory system functions to extract information about objects and events from properties of the soundwaves that it detects. It detects and groups the frequency components of the soundwave that reaches the ears, from which it extracts information about the object and the events that caused the object to vibrate. The frequency component groupings function to track or represent properties of the soundwave from which information about the source is extracted.

One might question the claim that the grouping of frequency components functions to track properties of the soundwave that reaches the ears rather than properties of the objects that produced the soundwaves. That would suggest an alternative characterisation of the function of the auditory system: it functions to extract information about the ways in which objects vibrate from the soundwaves it detects. It detects and groups the frequency components of the soundwaves that reach the ears on the basis of which it forms a representation of objects' vibrations from which information is extracted about the objects and the events that caused them to vibrate. The frequency component groupings function to represent the source objects' vibrations – the distal vibrations – from which information about the source is extracted. In both cases the end result is the same – information about objects is extracted – and in both cases our experience of sounds is determined by the way detected frequency components are grouped, but the two characterisations differ in what they imply about the connection between our experience of sounds and the sounds of which they are experiences.

If the frequency component groupings function to track objects' vibrations – the distal vibrations – then it is plausible that the sounds that our experience represents supervene on objects' vibrations. In that case, although our experiences of sounds would be determined by a proximal cause – the vibration of the soundwave that reaches the ears – what they *represent* would be determined by their distal cause – the vibrations of objects. If, on the other hand, the grouping of frequency components doesn't function to track objects' vibrations but vibrations of the soundwave detected by the ears, then there would be no grounds for claiming that the experiences of sounds determined by such groupings represent vibrations of objects. In that case, the sounds our experience represents supervene on properties of the soundwave and both our experiences of sounds and what our experiences represent would be determined by the more proximal cause. The correct characterisation of the function of grouping therefore matters for what we say about sounds.

How should we understand the function of grouping – as a process that attempts to tell us about distal vibrations of objects or as a process that simply tells us about the vibrations of the soundwave? Grouping processes operate on the initial sensory representations in the auditory system. These representations are produced as a result of the soundwave stimulating different areas of the basilar membrane to produce a frequency dependent excitation of nerves whose frequency dependent structure is preserved in tonotopically organised areas of the auditory cortex. These representations are like a spectrogram of the soundwave, encoding information about the intensity, frequency, and time course of its vibration: they represent (in the sense of embodying information about) frequency components of the soundwave. Grouping processes operate on these representations with the result that subsequent processes treat various represented components as belonging together. We shouldn't think of a grouping as a functioning to produce distinct representations of groups of frequency components. To say that components are grouped is simply to say that subsequent processes treat represented frequency components as a group.[42]

Should we think of groupings as representations of the distal vibration? Is grouping components together sufficient for those groups to function as representations of the distal cause? Since grouping treats together only those components that are likely to have been produced by the same object, groups of components will normally match components of the object's vibration; a consequence of grouping, therefore, is that a group of frequency components normally matches or corresponds to their distal cause – the vibration of an object. However, it doesn't follow from the normal correspondence of groupings with object vibrations that

[42] I am using the term 'grouped' are the name of a *process*, not the product of a process.

grouping is *functioning* to track objects' vibrations rather than vibrations of the soundwave: the mere fact of correspondence isn't sufficient for groupings to represent distal vibrations.

If detected frequency components are grouped *because* they are all part of the same distal vibration – because they co-occur distally – then we would have a reason to say that grouping functions to track the distal vibration. But frequency components are not grouped because they co-occur distally; they are grouped because they are produced by a single object and carry information about that object. Although it is a consequence of being produced by a single object that frequency components co-occur distally, that doesn't make distal co-occurrence itself explanatory of grouping. An explanation of why the auditory system groups harmonics that appeals simply to the fact that harmonics normally co-occur, for example, wouldn't tell us why harmonics normally co-occur or why their co-occurrence matters to the auditory system. Grouping is a process that treats components together because they are likely to have been produced by the same object and carry information about that object, and not because they are likely to have co-occurred: frequency components are *not* grouped by the auditory system *because* they are part of the same distal vibration.

The alternative characterisation of the function of the auditory system is therefore mistaken. We shouldn't think of grouping as an attempt to recover the distal vibration from the proximal vibration, and we cannot understand the function of the auditory system in terms of the distal vibration and without reference to the source object. The auditory system does not function to tell us about the way objects are vibrating. It *uses* the fact that the way an object vibrates carries information about the object and structures a soundwave to enable us to perceive that object and its properties. Information about the object can be extracted directly from properties of the *soundwave* by a process that involves auditory grouping, no part of which requires the auditory system to represent how the *object* is actually vibrating. Since information about the object can be extracted directly from the soundwave, it would be functionally pointless for the auditory system to use the way a soundwave vibrates as a guide to how the object that produced it vibrates only then to extract information about the object from that vibration. Thus our experience of sounds depends functionally on patterns instantiated by the soundwave rather than by the source of the sound, and it is implausible to think that the experience of sounds represents patterns instantiated by the source, rather than patterns instantiated by the soundwave which may sometimes also be instantiated by the source.

Our experience of sounds contrasts with our visual experience of objects and their properties in a significant respect. Visual experience represents the surface reflectance of objects – the perceived lightness of an object's surface. The amount of light that is reflected by the surface of an object depends on both the level of illumination of the surface and its reflectance. A light surface dimly illuminated can reflect the same light that a brightly illuminated dark surface reflects; since the dark surface can still be perceived as darker than the light surface, the lightness of an object is not determined directly from the proximal stimulus – the intensity of light that reaches the eyes. In order to accurately represent the object's surface reflectance property the visual system must compensate for changes in illumination.[43]

Although the auditory system functions to correctly represent other properties of objects, it doesn't function to correctly represent objects' vibrations. A good example is provided by loudness. How loud a sound is experienced to be is (approximately) determined by the amplitude of its associated frequency components.[44] The amplitude of the frequency components of the soundwave detected by the ears is determined by the amplitude of the object's vibration together with the distance of the object from the perceiver (amplitude decreases with distance). If the auditory system were functioning to represent the object's vibration then we would expect it to compensate for the distance of the object from the perceiver; only then would it correctly represent the amplitude of the object's vibration. It doesn't do so. The loudness of the sounds we hear is determined by the amplitude of the soundwave that affects the ears and not by the amplitude of the vibration of the object that produced the soundwave. Two objects vibrating at different amplitudes produce sounds which are experienced as equally loud for as long as the objects are at different distances from the perceiver. There is no auditory equivalent to the lightness

---

[43] Despite over a century of sustained investigation, how the perceptual system does this is not fully understood. For an excellent survey and discussion see Gilchrist 2006.

[44] How loud a sound is experienced to be should be distinguished from how forceful or violent the event that produced the sound is experienced to be. Changing the apparent loudness of the sound made by a stick striking a drum doesn't change how hard – with what force – the stick is heard to strike the drum.

constancy processes in vision.[45] This further supports my suggestion that our experience of sounds represents patterns instantiated by the soundwave rather than by objects.

The conclusion that our experiences of sounds can be veridical even when those sounds don't correspond to their sources seems better to capture our intuitions about the veridicality of our experience than the alternative. When we have an experience of sounds produced by loudspeakers, for example, we have an experience of sounds that don't correspond to their sources. According to the objection, such an experience misrepresents something as instantiating a pattern of frequency components: it misrepresents an object as instantiating the sound, when no such object exists. According to the objection, then, loudspeakers produce experiences of merely apparent sounds – of sounds that don't exist. That strikes me as a very counterintuitive conclusion, one that threatens the veridicality of many of our everyday auditory experiences.

Many of the everyday sounds we experience have been distorted or changed in some way during their transmission. The sounds of cars passing outside, for example, or of a voice from next door, are altered during their transmission through the structure of the building. Various frequency components of the soundwaves these objects produce are filtered or altered, so that the sounds we experience sound different to the way they would have sounded had they not been filtered. Since the pattern of frequency components instantiated by the soundwave that produces these experiences is different to that instantiated by the sources the sounds, any account of sounds that identifies them with properties of their sources would have to say that these experiences of sounds are not veridical.[46] On such an account many, if not

[45] For this reason, my argument that sounds supervene on the proximal cause of our experiences of sounds doesn't carry over to the case of colours: my argument does *not* have the implication that colours supervene on properties of the light that affects the eyes.

[46] Robert Pasnau has argued that any account, such as mine, according to which sounds supervene on soundwaves is incompatible with viewing our auditory experience as generally veridical on the grounds that we hear sounds 'as being at the place where they are generated' (1999, p.311). On my view it is the sources of sounds that we hear as located, not the sounds they produce. Pasnau's examples of hearing the location of sounds are for the most part examples of hearing the location of the sources of sounds. He rejects the suggestion that we hear the location the sources of sounds in virtue of hearing sounds that are spatially distinct from them as 'odd' on the grounds that it makes hearing 'indirect in the most unlikely way' (p.318). It follows from my account of the relation between sounds and their

most, of our normally produced auditory experiences turn out to be non-veridical. Such a consequence is usually thought to be a decisive reason for rejecting any putative account of perception.

I've said something about the correctness conditions for an experience of a sound. An account of the correctness conditions of our experience of sounds doesn't settle all questions about their identity conditions. This focus of this paper is auditory perception rather than sounds themselves – a detailed discussion of sounds is a topic for another paper – but a few brief remarks are in order. I have argued that an experience of a sound represents a pattern or structure instantiated by a soundwave. Does that mean we should identify particular sounds with instantiations of a type of pattern or structure?

Our normal ways of individuating sounds allows that two people, in different – even very distant – places, can hear the same particular sound – you and I both hear the same sound when we hear the sound of a gunshot. To deny this would be to allow that a single event – a gunshot, say – produces more than one sound: a sound heard by me, and a sound heard by you at a distance. Since people at different places who hear the same sound are not – or need not be – affected by the same instantiation of frequency components we cannot identify particular sounds with *instances* of a pattern or structural type. Similarly, our normal ways of individuating sounds allows that two people hear the same sound even if they hear it as having different qualities. The sound of a gunshot heard close by may be different – louder, sharper – to that same sound heard at a distance; it is, nonetheless, the same sound. But again, since the instantiation of frequency components must be different in the two cases, and may even be an instance of a different pattern, the sound cannot be identical to an instance of a pattern type.

If sounds are not identical to instances of pattern types, then could they be the pattern types themselves? Our normal way of individuating them treats sounds as particular things such that we can allow that two sounds may be qualitatively the same – the same type of sound – and yet be distinct individual sounds. Two sounds that are indistinguishable, for example, are usually counted as distinct if they are produced by

---

sources that sounds are not always located at their sources and that our auditory experience is generally veridical. I discuss the spatial content of auditory experience in more detail in (****).

different sources. Thinking of sounds as pattern types doesn't allow us to make this distinction. Furthermore, if sounds are things that come into being when they are produced then for any sound there is a time before which it did not exist, a time at which it came into existence and, presumably, a time at which it will cease to exist; although instances of pattern types have these temporal properties, pattern types themselves do not.

Any account of sounds should, as far as possible, accommodate our normal ways of individuating sounds. The ontological category that comes closest to doing so is that of particularized types or abstract individuals: to view sounds as abstract individuals would be to view them as belonging to the same ontological category as symphonies and other multiply instantiated art works, or to the same category as words (on Kaplan's account of the ontology of words).[47] To claim that sounds are abstract individuals is not, of course, to deny that sounds are instantiated by soundwaves any more than to claim that words are abstract individuals is to deny that words are instantiated by, for example, patterns of ink on paper. It simply allows the possibility that a sound, like a word, may be instantiated at more than one place and time.

Viewing sounds as abstract individuals is consistent with the plausible view that a recording of a sound, when played, can reproduce the very same sound as was originally recorded, so that, in hearing a recording, we hear the very same sound again. It is in virtue of that, I suggest, that in hearing a recording of Winston Churchill's voice we hear Winston Churchill; it explains, too, the sense we have that hearing a recording of a person brings us into closer contact with them than seeing a photograph of them does.[48]

10. Perception dependence

[47] This idea that sounds are abstract individuals was suggested to me by Mike Martin. For Kaplan's account of the ontology of words, see Kaplan (1990).
[48] But see Walton's discussion of the transparency of photographs in Walton (1984).

I have argued that our experiences of sounds are produced by a process whose function is to produce experiences of sounds that correspond to their sources: when it is successful, the sounds we experience correspond to their sources; when it is not successful, we experience sounds that do not correspond to their sources.

The principles that determine how the auditory system groups frequency components to produce experiences of distinct sounds – and so which explain why we hear the sounds we hear – can only be understood relative to the auditory system's function. We cannot give any account or explanation of what sounds are independently of such a process. The account I have given of the function of auditory perception has the consequence, therefore, that it is not possible to say what sounds are instantiated by a soundwave independently of the auditory system's capacity to detect them. In that sense, what sounds there are depends on what sounds we would experience there to be. Sounds are perception dependent.

Being perception dependent in this way does not mean being subjective. The claim that we cannot give any account of sounds independently of our capacity to perceive them does not imply that individual sounds are experience or mind-dependent, or otherwise subjective. I am not, for example, claiming that sounds are mental objects, that they are analogous to visual sense data. Visual sense data, as they are usually conceived, are mind-dependent objects of experience: it is supposed that having an experience as of an object is sufficient for the existence of an object – a sense datum – of which one is aware. Nothing in my account entails that having an experience of a sound is sufficient for the existence of that sound.

On my view, our experience of sounds represents the existence of an instantiated pattern or structure of frequency components and so has correctness conditions – we can make sense of the experience being veridical or non-veridical. Sounds are patterns or structures instantiated by soundwaves, and the instantiation of a pattern or structure by a soundwave is a perfectly objective – subject independent – matter. The fact that a soundwave instantiates a particular sound, therefore, is a fact about the soundwave that obtains independently of anyone having an experience of the sound, and having an experience as of a sound is not sufficient for that sound to exist. Which of the patterns of frequency components instantiated by a soundwave are sounds, however, is determined by how the auditory system would group those components. What sounds are instantiated by a soundwave therefore depends on the capacity of the auditory system to perceive them.

I began by describing the common view of auditory perception as the view that auditory perception functions to tell us about sounds and their properties. Whilst there's a perfectly good sense in which my account is an account of the perception of sounds and their properties, my account differs from the common view in two important respects. First, what sounds there are is perception dependent in the way I have just described. Second, our capacity to perceive sounds is part of a perceptual process whose function is to perceive the *sources* of sounds. The capacity to perceive sounds depends on the capacity to perceive the sources of sounds and not, as the common view has it, the other way around.

From an evolutionary point of view it makes good sense that auditory experience should represent the sources of sounds since it is the sources of sounds, rather than the sounds themselves, that have an impact on our survival and prosperity. Sounds, in contrast, are causally insignificant. It is puzzling, then, why we are aware of sounds as well as their sources: it is difficult to imagine what evolutionary advantage an awareness of sounds could confer on an animal that was already capable of perceiving the sources of sounds. It is a puzzle to which I don't have a solution.[49]

11. Content of auditory experience

The auditory system functions to represent sounds 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 represents both sounds and the sources of sounds, and we normally experience sounds that correspond to their sources.

How do we experience the connection between the sounds we experience and the sources of those sounds? When we have an experience of a sound and its source we are not independently aware of two objects as we are, for example, when we have a visual experience of two marks on a piece of paper. In the visual case we could – by covering one of the marks or shifting our attention – be aware of either mark without being aware of the other. This isn't true of our awareness of a sound and its source.

[49] One might speculate that an awareness of sounds rather than their sources plays an essential role in communication and it is this which conveys evolutionary advantage; though the implausibility of psychoacoustic accounts of speech perception over motor or direct-perception accounts of speech perception might suggest otherwise. See Massaro 1998, and Liberman, 1996.

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.

It is important to distinguish between the claim that we hear sounds as *appearing to have* the property of being produced by a source and the claim that we hear sounds *as having* the property of appearing to have been produced by a source. We can describe the character of a sound – the way the sound appears to us – in terms of its intrinsic auditory qualities of pitch, timbre, and so on. We cannot describe our experience of a sound as apparently having been produced by a source of a certain kind simply in terms of those auditory qualities of the sound that determine how that sound appears. This is clear in the case of our experience of the apparent location of the source of a sound. Two sounds that are otherwise indistinguishable can seem to come from sources located at different places; in hearing these two sounds we hear them as having been produced by sources with different (spatial) properties without hearing any difference in the auditory qualities of the sounds. In general, hearing a sound as seeming to have a source of a certain kind is not a matter of hearing it as having certain auditory qualities.   Therefore, hearing a sound as produced by a source is not simply a matter of hearing a sound as *appearing* some way – as having certain intrinsic properties.[50]

---

[50] The suggestion that sounds appear to have certain sources might seem plausible given that frequency component groupings both carry information about what produced them and determine what sounds we hear: we might be tempted to identify sounds with frequency component groupings and so think of sounds as themselves bearers of information. To do so would be mistaken. Whilst it is true that information is carried by soundwaves, it is extracted by the auditory system as part of the process that

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. 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.

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 other than sounds that surround-sound systems in the cinema are so effective.

---

produces our experiences of sounds and their sources, we are not aware of the relationships amongst frequency components in virtue of which they carry information and we don't experience sounds as having features that indicate what produced them.

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.[51]

The claim that auditory experience represents sounds as having been produced by their sources might seem puzzling: the sources of sounds are objects – how can *auditory* experience represent objects? It 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.[52] 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 in this way 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?[53] We see tomatoes *as* tomatoes and not simply as objects having a tomato-like appearance,[54] but how can visual experience represent something as having more than the *appearance* of a tomato – as having a

---

[51] The illusion shows the immunity to judgment that is characteristic of is the content of an experience as opposed to the content of judgement.

[52] 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.

[53] 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

[54] At least it is arguable that we do

certain colour, shape, and so on – how can it represent it as having whatever property it is (presumably a certain genetic structure) that determines whether something is a tomato? In representing something as cubic our visual experience represents it as having properties that go beyond the properties that actually determine how it appears. In representing something as a tomato our visual experience represents it as having properties that go beyond those that *could* determine how it appears.

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 visual experience is also true of auditory experience, and whatever explanation we give of how visual experience can have content that represents material objects will also apply to auditory experience. Therefore, the claim that auditory experience represents sounds as having been produced by their sources is no more puzzling or problematic – and so no more objectionable – than the claim that visual experience represents objects as material objects.

Although a similar explanation can be given both of how visual experience and auditory experience represents objects, visual and auditory experiences do not represent objects in exactly the same way. For example, in seeing a bowling ball rolling down a bowling lane we normally see it as having a range of properties: we see its colour, its shape and size, its location relative to other objects, the surface on which it rolls, and so on. We see these properties of the ball and its surroundings in virtue of our visual experience representing them. In hearing a bowling ball rolling down a bowling lane it is possible to hear it as such; in doing so we have an auditory experience as of an object rolling along a hard and smooth surface, an experience that wouldn't be veridical if it wasn't produced by an object rolling on a hard surface. If we took this experience at face value we would judge that there is an object rolling on a hard surface. But when we hear a ball as rolling our experience represents far less about the ball and its surroundings than the visual experience, and what it does represent it may represent in a way that is less determinate than the way visual experience would.

When we visually experience a ball as rolling the ball looks or appears spherical – the experience represents the ball as spherical. When we have an auditory experience of a ball as rolling the ball doesn't appear spherical. The auditory system is simply not capable of detecting the geometrical properties of objects with any precision; therefore, in representing a ball as rolling our auditory experience does not represent it as having a determinate shape – as being spherical rather than cylindrical, for example. But having a shape is not simply a matter of having a certain appearance: the shape of an object has implications for how it will behave in its interactions with other objects – it helps to determine the object's causal powers.[55] Although the auditory system cannot detect geometrical properties with any precision it can detect the casual interactions of objects including, for example, that an object is rolling on a hard surface. So although auditory experience may not represent objects as having a geometrical shape, it may represent them as having the causal properties that govern their interactions with other objects. A ball rolls in virtue of being spherical. In hearing a ball as rolling our auditory experience doesn't represent it as spherical but as having a causal property: the property shared by all objects that have a tendency to roll. That property is a determinable whose determinates are the shapes – cylindrical, spherical, and so on – that enable objects to roll.

Something similar is plausibly true of visual experience. In representing an object as having a shape, our visual experience does not simply represent it as having

[55] See Campbell 2005, pp.216 ff. This paragraph and the next draw on Campbell's discussion of the relations between tactile and visual perception of shape.

a certain appearance, but as having a causal property that partly determines its interactions with other objects. When we visually perceive the shape of an object we perceive its causal significance: in seeing an object as spherical we see it as having a tendency to roll. If that's right, then we can think of both auditory and visual perception as capable of representing the causal powers an object has in virtue of having the shape it has – we can both hear and see that an object has a tendency to roll – but only in the case of vision do we also experience the object as having a geometrical appearance.

Although the examples I have described show that our auditory experience can represent a sound as having been produced by an object, it does not follow that our auditory experience always represents a sound as having been produced by an object, and although the examples I have describe suggest that our auditory experience can represent the sources of sounds as having a variety of different properties, it doesn't follow that it always represents them as having such properties. Although an experience may represent the location of the source of a sound, for example, we have experiences which don't do so: we sometimes hear a sound and cannot tell where it comes from. This is another way in which auditory experience differs from visual experience. Normally, when one sees something one experiences it as being determinately some way for every way it is possible to visually experience something to be. It would be unusual or abnormal to have a visual experience of an object that didn't represent it as being coloured, of a certain size and shape, at a certain location, and so on. The same is not true of auditory experiences. Our auditory experience of the source of a sound may be in many respects indeterminate and tell us little about the nature of the source.

If experience represents sounds as having been produced by their sources it must represent causal relations. We are familiar with experiences of seeing one event cause another, as Peacocke says:

> anyone who sees the child's hand knocking over the tower of blocks, or a fork-lift truck as lifting a crate, has [experiences as of one event causing another]. These experiences would not be adequately characterised as seeing an event of one type following an event of another type. Rather, taking the experiences at face value, one would be disposed to judge that the child's movement caused the tower to fall

> over or to judge that rising of the fork-lift truck's arms caused the crate to go up (1986, p.156).

In these examples, apparent causation between two events is visually perceived. I am suggesting that, just as visual experience can represent a causal relation between two objects, so auditory experience can represent a causal relation between a sound the thing that produced it. Of course, auditory experience doesn't represent a causal relation between two objects, but between an object and a sound; nonetheless, in experiencing a sound as produced by a source, our experience is representing an instance of a relation of the same kind as our visual experience represents in the cases Peacocke describes. Thus, taking an experience of hearing the sound of a dog's bark at face value, one would be disposed to judge that a dog is responsible for the sound that one hears, that the dog's barking is *causing* or *producing* that sound.[56]

Someone might still object to my description of these experiences as experiences as of causation for the reason that we can never have such experiences.[57] Our concept of causation, it might be argued, is the concept of a kind of relation which we could not simply perceive to be instantiated. Peter Menzies, for example, suggests that the counterfactuals involved in an instance of causation make it a relation that 'cannot plausibly be claimed to be an object of direct awareness' on the grounds that the truth of a counterfactual cannot be perceived (1993, pp. 202-3). The concept of a cause to which I am appealing, however, is one of a whole range of causal concepts that feature in our everyday thought and language, concepts which

---

[56] Of course, I am not suggesting that the circumstances in which in the auditory system represents two objects as causally related are the same as those in which the visual system represents sounds and their sources as casually related. For the claim with respect to vision, see Michotte (1963, esp. appendix 2); and see Bruce and Green (1990, p.333) for a discussion and other references. Bruce and Green are sceptical of Michotte's claim that causality is *directly* perceived, but not of the claim that we do have experiences of the sort described by Michotte. It's just that they think that an explanation of our experience's representing causality must appeal to computations or inferences performed by the visual system.

[57] In his account of the content of visual experience John Searle (1983, ch.2) argued that the content of a visual experience of a truck is veridical only if the experience is caused by the truck. His argument has been widely criticized on the grounds that it is implausible to think experience represents such a causal relation. What is implausible about Searle's view is not the claim that visual experience represents a causal relation, but that it represents a reflexive causal relation between the object of an experience and the experience of that object. The causal relation I claim to be represented in auditory experience is between objects – sounds and their sources – not objects and experiences, and so is not similarly implausible.

include: scrape, push, carry, knock over, squash, make and so on,[58] and not that provided by philosophical analysis. We should view a philosophical analysis of causation as giving an account the relation to which the everyday concept refers and not an account of the basis on which we apply the everyday concept. For as long as we allow that people possess and use such everyday causal concepts, and can apply them to things on the basis of perceiving the interactions between objects or, as in the case of auditory experience, on the basis of experiencing of sounds and their sources, then we should allow that causality, in this sense, can be perceived. People do possess and use such concepts; Anscombe is undoubtedly right when she says that

> as surely as we learned to call people by name or to report from seeing it that the cat was on the table, we also learned to report from having observed it that someone drank up the milk or that the dog made a funny noise or that things were cut or broken by whatever we saw cut or break them (1971, p.69).

In claiming that our auditory experience represents sounds as produced by their sources I am claiming no more than that our experience represents a relation such as these.[59]

Claiming that we experience sounds as produced by their sources might seem to go against those who claim that a pure sound world is conceivable.[60] In fact, it does not do so. My claim is not that sounds themselves require or entail the existence of something other than sounds – that sounds have the intrinsic property of having been produced by something – but that our experience – by representing sounds as produced by something represents sounds as related to something other than sounds. So although our auditory experience, if it is to be veridical, requires the existence of things other than sounds, there is nothing prima facie incoherent in the idea of a pure

[58] This list is from Anscombe (1971, pp. 68-9).
[59] Anscombe points out that the apparent perception of such things may only be apparent: we may be deceived by false appearances. It should be noted, too, that we can accept that we have such experiences of causation without committing ourselves to any particular account of the nature of causation in the world (c.f. Peacocke 1986, p.156).
[60] Famously Strawson (1959, ch.2); see also Scruton (1997, ch.1).

sound world, a world of sounds that were not produced by sources. (Perhaps a powerful deity could simply bring about soundwaves that instantiate patterns of frequency components.) If a pure sound world is conceivable, it involves conceiving of sounds as unheard, as existing independently of our experience of them.

12. Conclusion

I began by describing the pipe-organ illusion and asking in what way the experience produced by it is illusory. It follows from the account I have given of the function of the auditory system that the experience misleads us about the source of the sound we hear. In hearing the sound produced by the pipe-organ illusion, we have an experience of a sound that seems to have been produced by something that did not in fact produce it. Although our experience of the sound is veridical – there really is a sound that we hear and it is the way we hear it to be – our experience of the source is not. There are other auditory illusions and strange auditory phenomena that I have not had space to discuss.[61] Many of them occur as a consequence of the fact that we can artificially manipulate frequency components to produce sounds in an abnormal way (from the point of view of the auditory system) – in a way that they could never naturally be produced. All such illusions can be accommodated within the account that I have given. The pattern of explanation of these auditory illusions is perhaps unique to audition; they occur as a result of the fact that we perceive the sources of sounds by perceiving the sounds they produce. In that sense our auditory experience is mediated or indirect.

I have not had space to discuss alternative views of the nature of sounds, but the account I have given is inconsistent with a number of recent accounts. Those accounts have all argued, in one way or another, that sounds are properties of or events involving their sources.[62] They have done so on the grounds that we experience sounds as being located at their sources. On my account, sounds are distinct from their sources and – as the pipe-organ illusion shows – may not in fact be produced by a single source; on my account it is the *sources* of sounds that we

[61] Some of these are available on Diana Deutsch's CDs 'Musical Illusions and Paradoxes' and 'Phantom Words and Other Curiosities'.
[62] See Casati and Dokic (1994), Pasnau (1999), and O'Callaghan (2005).

experience to be located rather than the sounds that they produce. These alternative accounts all share the assumption that auditory perception functions in the same way as visual perception; it differs only in enabling the perception of sounds. One of my aims has been to show that auditory perception is different to visual perception and that an account of sounds cannot be given independently of an account of auditory perception.

Many further questions remain – about speech and music perception; about sound and sound-source recognition; about the detailed content, especially the spatial content, of auditory experience; and about the connections between auditory and visual experience. By giving an account of the function of auditory perception and of the content of auditory experience I hope to have provided a framework within which these questions may fruitfully be addressed.

References

Anscombe, Elizabeth. 1971. Causality and Determinism. In *Causation and Conditionals*, edited by E. Sosa. Oxford: Oxford University Press.

Blauert, Jens. 1997. *Spatial hearing: the psychophysics of human sound localization*. Rev. ed. Cambridge, Ma.: MIT Press.

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

Bruce, Vicki, and Patrick Green. 1990. *Visual Perception: Physiology, Psychology and Ecology*. London: Lawrence Erlbaum Associates.

Butler, R. J. 1965. *Analytical philosophy: Second Series*. Oxford: Basil Blackwell.

Cabe, P., and J. B. Pittenger. 2000. Human sensitivity to acoustic information from vessel filling. *Journal of* 9:211-214.

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.

Casati, Roberto, and Jérôme Dokic. 1994. *La philosophie du son*. Nimes: Editions Jacqueline Chambon.

Chowning, John. 1999. Perceptual fusion and auditory perspective. In *Music, Cognition, and Computerised Sound,* edited by P. R. Cook. Cambridge, Ma.: MIT Press.

DeBellis, Mark. 1995. *Music and conceptualization*. Cambridge: Cambridge University Press.

Deutsch, D. 1999. Grouping mechanisms in music. In *The psychology of music*, edited by D. Deutsch: Academic Press.

Eilan, Naomi, Rosaleen A. McCarthy, and Bill Brewer. 1993. *Spatial representation: problems in philosophy and psychology*. Oxford: Blackwell Publishers.

Evans, Gareth. 1982. *The Varieties of Reference*. Edited by J. McDowell. Oxford: Clarendon Press.

Fletcher, Neville H., and Thomas D. Rossing. 1998. *The Physics of Musical Instruments*. 2nd ed. New York: Springer.

Fowler, C. A. 1991. Auditory perception is not special: We see the world, we feel the world, we hear the world. *Journal of the Acoustical Society of America* 89:2910-2915.

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. 1993a. How do we hear in the world? Explorations in ecological acoustics. *Ecological Psychology* 5:285-313.

———. 1993b. What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology* 5:1-29.

Gelfand, Stanley A. 1998. *Hearing: an introduction to psychological and physiological acoustics*. 3rd ed. New York: M. Dekker.

Gilchrist, Alan. 2006. *Seeing Black and White*. New York: Oxford University Press.

Griffiths, T. D. 2002a. Central auditory pathologies. *British Medical Bulletin* 63:107-20.

———. 2002b. Central auditory processing disorders. *Current Opinion in Neurology* 15:31-3.

Griffiths, T. D., Christian Büchel, Richard S.J. Frackowiak, and Roy D. Patterson. 1998. Analysis of temporal structure in sound by the human brain. *Nature Neuroscience* 1 (5):422-7.

Griffiths, T. D., Rees A., and Green G. G. R. 1999. Disorders of human complex sound processing. *Neurocase* 5:365-378.

Kaplan, David. 1990. Words. *Proceedings of the Aristotelian Society Supplementary Volume* LXIV.

Kosslyn, Stephen M., and Daniel N. Osherson. 1995. *Visual Cognition*. 2nd ed. Vol. 2, *An Invitation to Cognitive Science*. Cambridge, Ma.: MIT Press.

Kunkler-Peck, A., and M. T. Turvey. 2000. Hearing shape. *Journal of Experimental Psychology: Human Perception and Performance* 1:279-294.

Lakatos, S., S. McAdams, and R. Caussé. 1997. The representation of auditory source characteristics: Simple geometric form. *Perception and Psychophysics* 59:1180-1190.

Li, X., R. J. Logan, and R. E. Pastore. 1991. Perception of acoustic source characteristics: Walking sounds. *Journal of the Acoustical Society of America* 90:3036-3049.

Liberman, A. M. 1996. *Speech: A special code*. Cambridge, MA: The MIT Press.

Lutfi, R. A., and E. Oh. 1997. Auditory discrimination of material changes in a struck-clamped bar. *Journal of the Acoustical Society of America* 102:3647-3656.

Massaro, Dominic W. 1998. *Perceiving talking faces: from speech perception to a behavioral principle*, *MIT Press/Bradford Books series in cognitive psychology*. Cambridge, Ma.: MIT 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.

McAdams, Stephen, and Emmanuel Bigand, eds. 1993. *Thinking in Sound*. Oxford: Oxford University Press.

McBeath, Micheal K, and John G Neuhoff. 2002. The Doppler effect is not what you think it is: Dramatic pitch change due to dynamic intensity change. *Psychonomic Bulletin and Review* 9 (2):306-313.

Menzies, Peter. 1993. Laws of Nature, Modality and Humean Supervenience. In *Ontology, Causality and Mind*, edited by J. Bacon, K. Campbell and L. Rheinhardt. Cambridge: Cambridge University Press.

Michotte, A. 1963. *The Perception of Causality*. Translated by P. Heath. New York: Basic Books.

Nakayama, K., N. H. He, and S. Shimojo. 1995. Visual Surface Representation: A Critical Link between Lower-level and Higher-level Vision. In *Visual Cognition*, edited by S. M. Kosslyn and D. N. Osherson. Cambridge, Ma.: MIT Press.

Neuhoff, John. 2004. Auditory motion and localisation. In *Ecological Acoustics*, edited by J. Neuhoff. London: Academic Press.

O'Callaghan, Casey. 2005. Sounds and Events.

O'Shaughnessy, Brian. 1957a. An impossible auditory experience. *Proceedings of the Aristotelian Society*. 57:53-82.

———. 1957b. The location of sound. *Mind* 66:471-490.

———. 1971-1972. Processes. *Proceedings of the Aristotelian Society*. 72.

Pasnau, Robert. 1999. What is Sound. *Philosophical Quarterly* 49:309-324.

———. 2000. Sensible Qualities: The Case of Sound. *Journal of the History of Philosophy* 38:27-40.

Peacocke, C. 1982. *Sense and Content*. Oxford: Clarendon Press.

———. 1986. *Thoughts: An Essay on Content*. Oxford: Basil Blackwell.

———. 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.

Repp, B. H. 1987. The sound of two hands clapping: An exploratory study. *Journal of the Acoustical Society of America* 81:1100-1110.

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.

Searle, John. 1983. *Intentionality*. Cambridge: Cambridge University Press.

Strawson, P. F. 1959. *Individuals*. London: Methuen.

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).

Walton, Kendall L. 1984. Transparent Pictures: On the Nature of Photographic Realism. *Critical Inquiry* 11:246-277.

Warren, Richard M. 1999. *Auditory perception: a new analysis and synthesis*. Cambridge: Cambridge University Press.

Warren, W. H., & Whang, S. 1987. Visual guidance of walking through apertures: Body-scaled information for affordances. *Journal of Experimental Psychology: Human Perception and Performance* 13:371-383.

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.

Draft 2/2/2007

Matthew Nudds