

Restricted Auditory Aspatialism

Douglas C. Wadle

Abstract

Some philosophers have argued that we do not hear sounds as located in the environment. Others have objected that this straightforwardly contradicts the phenomenology of auditory experience. And from this they draw metaphysical conclusions about the nature of sounds—that they are events or properties of vibrating surfaces rather than waves or sensations. I argue that there is a minimal, but recognizable, notion of audition to which this phenomenal objection does not apply. While this notion doesn't correspond to our ordinary notion of auditory experience, it does—in conjunction with our lack of an uncontroversial individuation of the senses and recent interest in distinctively multisensory features of perceptual experiences—raise the possibility of more expansive notions of audition, including some that do plausibly count as corresponding to our everyday notion of audition, that lack the spatial phenomenology cited in the objection. Until this possibility is ruled out, the phenomenal objection and metaphysical conclusions drawn from it remain inconclusive.

- 1 *Introduction*
- 2 *Setting the Target: Restricted Auditory Aspatialism*
 - 2.1 *Strawson's restricted auditory aspatialism*
 - 2.2 *A new motivation for restricted auditory aspatialism*
 - 2.3 *The target notion of a purely auditory experience*
- 3 *Spatial Hearing*
 - 3.1 *Monaural cues*
 - 3.2 *Binaural cues*
- 4 *RAA, unrestricted aspatialism, and the ontology of sound*
- 5 *Conclusion*

1 Introduction

Frequently, on nighttime walks through my neighbourhood, I hear a tone—a bit like the whistle of a tea kettle but lower pitched, not loud but not faint—in the distance. I call it the Whistler. I have no idea what causes the Whistler, but it sounds as though it is rather far away and so must be fairly loud at its source. It comes from the Southwest. Were you to come walking with me, I trust you would corroborate all of this and that you would find your ability to do so as unremarkable as I find my ability to have made these observations in the first place. Nevertheless, some philosophers have

argued that we do not enjoy spatial audition—that is, we do not hear sounds as located in the environment, even in the bare sense of merely off in a given direction or at some particular distance (Machlachlan [1989]; Malpas [1965]; Nudds [2001], [2009]; O’Shaughnessey [2000], [2009]; and, in a qualified form discussed below, Strawson [1959]). Call this view ‘auditory aspatialism’. In spite of my evening encounters with the Whistler, I hereby submit this as my application for (qualified) membership in the aspatialist club.

In the first half of this paper I provide the philosophical background and motivation for my approach to auditory aspatialism. In the second half, I provide empirically-based arguments in its favour. These arguments consider the receptivity of the primary auditory receptor—the basilar membrane—to features of a sound wave impinging upon it and the role of these features in determining spatial features of the sound or its source. I then show how these results undermine metaphysical conclusions about the nature of sounds that the phenomenal objection has been thought to support.

2 Setting the Target: Restricted Auditory Aspatialism

Auditory aspatialism confronts an immediate and obvious worry, already evident in my description of the Whistler: We do experience sounds as at a distance in some particular direction. Aspatialism straightforwardly contradicts this phenomenological evidence. Auditory spatialists have argued, from this fact and the assumption that our perceptual experiences don’t systematically mislead us, that sounds are events or properties of their sources (Casati and Dokic [2005], [2009]; O’Callaghan [2007], [2010]). Recent aspatialists have responded to the objection by denying that we ever experience sounds as located and offering an error theory that explains away the recalcitrant phenomenology—generally in support of the view that sounds are (instantiated by) pressure waves (Nudds [2009]; O’Shaughnessey [2000], [2009]). If the wave view is correct, then our auditory experience cannot present sounds as located in some particular direction at a particular distance without systematically misleading us because the sound wave isn’t located at a distance when heard. It is located right where the perceiver is.¹ Hence the need for an error theory.

Aspatialists can also respond to the objection by restricting their aspatialism to a special notion of audition to which the phenomenal objection does not apply. This is the course taken by Strawson ([1959]) in his original formulation of auditory aspatialism. Strawson’s aspatialism grew out of his project of descriptive metaphysics—an exploration of the relations between the concepts of space and objectivity in our conceptual system—rather than a concern with the metaphysics of sound

¹ Aspatialism is about the experience of sounds, not the sounds themselves. Proponents of the wave view don’t deny that sounds (waves, according to them) are located. They deny that we experience them as located—but see (Sorensen [2008], Chapter 14) for a defence of a distal wave view). Also note, both spatialists and aspatialists agree that perception does not systematically mislead us – at least not so dramatically as to mis-locate its objects to the degree necessary were one to embrace the wave view and accept that sounds are experienced as reported in the phenomenal objection.

(§2.1). I offer a different motivation for taking the restricted aspatialist approach; namely, that, when properly developed, restricted auditory aspatialism opens a new line of resistance against the phenomenal objection to unrestricted aspatialism—one that denies that the localization phenomenology cited in the objection is due to audition (§2.2).

The rough idea is this: If restricted auditory aspatialism is true, then there is some notion of audition to which the phenomenal objection does not apply, even if it isn't the intended target of the objection. This raises the possibility that there are characterizations of the ordinary (unrestricted) auditory experience to which the objection does not apply. This possibility becomes especially salient given the lack of any generally agreed upon individuation of the senses—or even that there is just one acceptable individuation of the senses—and recent interest in the possibility of features of perceptual experiences that aren't properly associated with any of the five standard senses but are constitutively multisensory.² Until the proponent of the phenomenal objection rules out this possibility, the phenomenal objection to unrestricted aspatialism remains inconclusive. (Problems for meeting these burdens will also be discussed in §2.2.) Furthermore, if the localization phenomenology isn't due to audition, and sounds are the objects of audition, then the metaphysical conclusions drawn from the phenomenal objection don't follow.

Of course, this all hangs on the truth of restricted auditory aspatialism. To assess restricted aspatialism, we need to clearly characterize the relevant restricted notion of audition. And we need a principled way of assigning features of a perceptual experience to audition. This is done in terms of components of perceptual mechanisms and the features of the stimulus to which they are sensitive (§2.3). These same tools can, I suggest, be extended in future work assessing the viability of this new line of resistance. I present the argument for the restricted form of auditory aspatialism in §3. I then return to the strategy for defending unrestricted aspatialism and how it undermines the metaphysical conclusions drawn from the phenomenal objection (§4). This should make the restricted aspatialist approach of interest to those concerned with the debate over the ontology of sounds, as well as more general questions concerning audition and spatial perception, without entangling us in Strawson's descriptive metaphysics.

2.1 Strawson's restricted auditory aspatialism

The trick, for the proponent of the restricted aspatialist approach, is to find the right notion of auditory experience—one that escapes the phenomenal objection while still being recognizably auditory. Strawson's restricted aspatialism provides a useful framework here. It is *the* historical

² On individuation of the senses, see (Coady [1974]; Fulkerson [2014b]; Gray [2013]; Grice [1962]; Heil [1983]; Keeley [2002]; Macpherson [2011b], [2011c], [2014]; Matthen [2015]; Nelkin [1990]; Nudds [2004]; Roxbee-Cox [1970]). On multisensory features, see (Bayne [2014]; Briscoe [2016], [2017], [2019]; Connolly [2014]; Fulkerson [2014a]; Macpherson [2011a]; O'Callaghan [2008], [2012], [2014a], [2014b], [2015], [2017a], [2017b], [2019]). See (Spence and Bayne [2014]) for a critical look at the evidence for multisensory experiences. See (Mandrigin [2021]; Wadle [2020]) for critical discussion of the arguments in favour of multisensory features.

articulation of the view. More importantly, discussing its ambiguities brings crucial desiderata for the target notion into focus and sets the stage for the new line of resistance mentioned above.

Strawson begins with the claim that, in an exclusively auditory experience, there would be no place for the application of spatial concepts (Strawson [1959], p. 65). He goes on to say that:

The fact that, with the variegated types of sense-experience which we in fact have, we can, as we say, 'on the strength of hearing alone' assign directions and distances to sounds, and things that emit or cause them, counts against this not at all. For this fact is sufficiently explained by the existence of correlations between the variations of which sound is intrinsically capable and other non-auditory features of our sense-experience. I do not mean that we first note these correlations and then make inductive inferences on the basis of such observation; nor even that we could on reflection give them as reasons for the assignments of distance and direction that we in fact make on the strength of hearing alone. To maintain either of these views would be to deny the full force of the words 'on the strength of hearing alone'; and I am quite prepared to concede their full force. I am simply maintaining the less extreme because less specific thesis that the *de facto* existence of such correlations is a necessary condition of our assigning distances and directions as we do on the strength of hearing alone. (Strawson [1959], p. 66)

In the passage above, Strawson embraces aspatialism with respect to both sounds and their sources.³ Despite this, he is explicit that his view is not prone to the phenomenological objection. He grants that we can locate sounds and sources 'on the strength of hearing alone' and insists that this is not a matter of making inductive inferences from the deliverances of auditory experience.⁴ Strawson can only maintain both positions if he draws a distinction between the sort of auditory experience relevant to the localization of sounds 'on the strength of hearing alone' and the sort of auditory experience relevant to his aspatialism—the purely auditory experience. This distinction is, for Strawson, to be drawn along the lines of information sharing with the other sensory modalities. So, how exactly do we break these inter-modal correlations in order to get at the purely auditory experience?

It will not be enough to simply subtract those experiences we associate with the other senses from our occurrent sensory experience. We are trying to drive a wedge between some specialized notion of auditory experience and the notion to which 'the full force' of the phrase 'on the strength of hearing alone' applies. What could the full force of this phrase be if not on the basis of one's auditory experience without the assistance, in experience, of the other senses (individuated in the common sense way)? But then simply subtracting the other senses from an occurrent experience won't give us

³ Contrast with Nudds's ([2009]) aspatialism, which concerns our experience of sounds but not their sources.

⁴ See (O'Shaughnessy [2000], [2009]) for a view on which sound localization is a matter of inference.

a specialized notion of audition. It will give us the ordinary notion of audition. So, unless Strawson is confused, this cannot be what he has in mind.⁵

Notice also that Strawson references correlations between experienced features of sounds and non-auditory features of sensory experience. Given that he does not think hearing the distance or direction of sounds or their sources 'on the strength of hearing alone' is a matter of noting these correlations and making inductive inferences from them, Strawson must be thinking of these correlations as operating subpersonally, prior to the ordinary auditory experience.⁶

There are two sorts of subpersonal correlation we must consider. The first is parasitic on previous conscious experience: Past awareness of the frequent coincidence of some spatial feature registered by a non-auditory modality with some feature of auditory experience leading to the automatic association of the given auditory feature with the given spatial feature. O'Callaghan's ([2017a]) interpretation of Strawson's purely auditory experience, as an auditory experience had by a subject who does not have and never has had any sensory experience in a modality other than audition, is an attempt to rule out just these sorts of correlations. On this interpretation, we are not having a purely auditory experience while lying on our backs in a sensory deprivation tank listening to music piped in through speakers on the ceiling of the tank because we have had experiences in other modalities prior to getting in the tank.

Suppose that learned associations between looming visual objects and increasing auditory intensity are necessary to experience a sound (source) as approaching due to a steadily increasing intensity. One of O'Callaghan's purely auditory experiences will not be able to avail itself of these learned correlations.⁷ But what about innate correlations?

O'Callaghan-style pure experiences do not rule out innate correlations to representations or capacities that could supply those contents lost by the isolation of the auditory experience from any historical interaction with the other senses. These representations or capacities might be associated with some other sensory modality or they might be premodal (in that they give input to sensory processes prior to their experiential output but do not belong to any one modality). For instance, changing intensity might be innately correlated to changing locations in a premodal egocentric

⁵ Despite this, this is one of the interpretations O'Callaghan ([2010]) gives him.

⁶ This is consistent with our being able to recognize the correlations and to make inferences from them or to offer them as reasons for our localization of the sound. But it doesn't require it.

⁷ We should also rule out learned correlations to non-sensory information, as well. This is generally passed over without remark, perhaps due to an assumption that any relevant non-sensory information will be, itself, derived—at least in part—from sensory information or to an assumption that perceptual content is nonconceptual.

coordinate system, allowing changes in intensity to carry information about distance. Or it might be innately correlated, via visual looming, to such a coordinate system associated with vision.⁸

So, if Strawson's aspatialism is to be assured of avoiding the phenomenal objection, it will need to isolate its purely auditory experience from any innate correlations to non-auditory resources, as well. Textual evidence indicates that Strawson was only concerned with learned correlations, supporting O'Callaghan's ([2017a]) interpretation. For instance, Strawson refers to 'correlations between the variations of which sound is intrinsically capable and other non-auditory features of our sense experience'. If such textual evidence is to be believed, this is where Strawson and I part company.⁹ My minimalism about the purely auditory experience cuts deeper than his. Mine excludes these innate correlations, too, because they are, after all, correlations to non-auditory resources and so shouldn't factor in an assessment of a purely auditory experience.

2.2 A new motivation for restricted auditory aspatialism

As a first step in motivating my approach to the purely auditory experience, consider the ventriloquism effect, in which visual input dominates the localization of an auditory stimulus. Is the spatial phenomenology associated with the ventriloquism effect auditory, visual, or audio-visual? If we say auditory, we implicate our eyes as part of the auditory mechanism, the stimulation of which gives rise to the experience. If we say visual or audio-visual, we risk making a great deal of what we ordinarily take to be spatial auditory phenomenology to be, in fact, non-auditory. (I have in mind all of the instances whereby our auditory localization is made more refined by visual input.¹⁰)

This is not to say that there is no way to resolve this tension with respect to the ventriloquism effect (or other cross-modal interactions). The point is that the choices we make about what is and isn't a part of the mechanism that generates our ordinary auditory experiences will impact which features of a perceptual experience we ought to class as auditory (and vice versa). But intuitions about which features of a perceptual experience ought to be classed as auditory come apart from intuitions about what is and isn't a part of our auditory mechanism. This tension must be resolved

⁸ There is empirical support for the claim that correlations between visual looming/receding and changes in intensity are innate, though this doesn't touch on the question of whether or not these correlations support perception of change in auditory distance from changes in intensity without concurrent visual stimulation (Orioli *et al.* [2018]).

⁹ It is not clear that we should read the evidence this way. After all, given the motivations for his aspatialism (his descriptive metaphysics project), he would likely happily omit innate correlations, too. Thankfully, we do not have to decide the exegetical question. Strawson has served his purpose for the present endeavour, and we can now set him aside.

¹⁰ Given that this is a difference in degree not kind, this will be so even if there are different mechanisms for resolving discrepant audio-visual inputs and for improving resolution of auditory spatialization by way of visual input. Whatever reasons we have for letting one of these mechanisms in will be reason enough to let the other in. Nor is it a problem that the ventriloquism effect and spatial recalibration seem to presuppose that (what is being classified as) auditory input carries spatial information of its own. Even if this is true, we haven't settled the question of whether audition, construed in this way, is free of other non-auditory inputs without which it couldn't have this information.

before we can say which features of our perceptual experiences ought to count as auditory—even (potentially) of our ordinary auditory experiences. Unfortunately, there is no uncontroversial proposal for individuating the standard five senses (sight, audition, smell, taste, and touch) and thereby determining which features of experience are due to them.¹¹ Perhaps we can be sure that some features are auditory (for instance, proper sensibles like pitch), but this certainty cannot be extended to features about which there is controversy (including sound localization phenomenology). But then we're not in a position to adjudicate the debate between aspatialists and spatialists about the ordinary auditory experience because we cannot yet say if the phenomenal features at play in the phenomenal objection are, in fact, properly auditory.

This possibility becomes even more salient in light of recent arguments in favour of the existence of distinctively multisensory features—features of perceptual experiences that are not associated with any of the standard five senses.¹² Whether or not there are such features certainly depends on how we individuate those senses, but—again—there is no generally accepted account of how to do this. So it remains a live possibility that auditory spatial phenomenology is one of these multisensory features.¹³ Even if we resort to sensory pluralism—the view that the individuation of a given sense is sensitive to the explanatory purposes to which it is put (Fulkerson [2014b]; Macpherson [2011c], [2014])—to alleviate this worry, we still run into difficulties. According to sensory pluralism, there are different ways of individuating audition depending on context. But then we need to know which context(s) is (are) relevant for assessing spatialism about audition. Until these issues are settled, the phenomenal objection is inconclusive. So, too, are the metaphysical claims it has been used to support: If it is unclear that audition presents its objects as located in the distal environment, we can't use this localization to draw conclusions about those objects (more on this in §4).

The way forward is to find some preferred way to demarcate the auditory mechanism and then to assess whether or not that mechanism could give rise to auditory experiences of sounds as

¹¹ See n.2 for references. While this literature tends to focus on cross-species individuation, the lessons apply within species as well. There is no generally accepted rule for resolving this tension revealed by the ventriloquism effect, for instance.

¹² See n.2 for references. Notice that O'Callaghan, himself, is a leading advocate for the existence of perceptual features not associated with any of the everyday senses, in addition to being an advocate of the phenomenal objection.

¹³ Generally, it is assumed in the multisensory features literature that auditory localization (including ventriloquism effect location phenomenology) is properly associated with audition, but no argument is offered for the intuition. Nor is it clear that this intuition can be reconciled with other feature association intuitions relied on in this literature (Wadle [2020]). But then it remains an open possibility that supposedly auditory localization phenomenology is a novel feature. Likewise, it remains possible that this phenomenology is properly associated with some other sense. We can be mistaken about such things—see (Spence [2016]).

located in a direction or at a distance.¹⁴ To do this we ought to start with the most minimal sensory mechanism built out of uncontroversially auditory components that will still count as auditory—the mechanism associated with the purely auditory experience—and see if it can deliver the relevant spatial properties. If it can, then no form of auditory aspatialism is true (restricted or not). If it can't, then restricted auditory aspatialism is true for that notion of audition. We can move from this result to results pertaining to the unrestricted aspatialist thesis, building upon the mechanism of the purely auditory experience until it does deliver these properties. Then we assess whether or not the mechanism specified should count as auditory.¹⁵ If it does, then the spatialist carries the day, provided it is the most restricted acceptable characterization of (ordinary) audition. If not, the phenomenal objection remains inconclusive (barring further refinements to what is and isn't an acceptable way of characterizing the mechanisms of ordinary audition).

We build out from the mechanism of the purely auditory experience by introducing capacities and informational input from other sources—both those associated with other sensory modalities and those that are not (innate premodal spatial representations, for example).¹⁶ Of course, we can always recover these spatial properties by positing such innate representations or capacities, but doing so comes with a burden: We need both a good reason to posit the existence (in us) of the representation or capacity and a good reason to think it part of the auditory mechanism—at least if this is going to secure the spatialist conclusion. And it seems unlikely that we will find good reason for thinking a given innate representation or capacity is best thought of as part of the auditory mechanism.¹⁷ Presumably the reason would be that it only interfaces with unequivocally auditory perceptual mechanisms. But given the many interactions of the sensory mechanisms, beginning at their early stages, it is not obvious that there will be an unequivocal specification of the auditory mechanism at the point we are likely to find such representations or capacities.¹⁸ This gives us further reason to

¹⁴ We need to ground the discussion in mechanisms because the dispute concerns whether some phenomenal features or representational contents should be attributed to audition, on its own. So we need to appeal to something other than phenomenology or contents to settle the question. The auditory mechanism is the obvious choice. More generally, focusing on sensory mechanisms to settle questions about phenomenology and content has the benefit of making these questions empirically tractable.

¹⁵ Assuming that we can settle on criteria for an acceptable characterization of ordinary audition. One need only look at the extensive literature on the individuation of the senses to realize that this is likely to be a difficult task.

¹⁶ On Bayesian accounts of perception, these resources will include priors. A complete Bayesian account will need to include a story about where the priors come from—whether acquired or innate. If we attend to the various inputs to minimal audition required to recover the features of an ordinary auditory experience—as I am urging—we will be able to tell which such inputs are learned from interactions with other senses and which are not.

¹⁷ Whether there is good reason to posit them in the first place (for any modality, not just audition) is a matter of some debate. For recent entries in that debate see (Cassam [2005], [2007]; Chomanski [2017]; French [2018]; Matthen [2014]; Schwenkler [2012]).

¹⁸ For a sampling of the empirical evidence for this claim, see (Calvert *et al.* [2004]; Driver and Noesselt [2008]; Ghazanfar and Schroeder [2006]; Schroeder and Foxe [2005]; Shimojo and Shams [2001]). Evidence specific to audition will be discussed below.

omit such representations or capacities (should they exist) from our unequivocally auditory starting point—they are not unequivocally auditory.

This paper carries out the beginnings of this project—it establishes that restricted aspatialism is true. Whether or not any or all acceptable characterizations of the mechanisms underwriting our everyday notion of audition capture the relevant spatial properties is left as a question for another time. But it is a question that I hope to make more tractable by showing (i) how we should go about characterizing the relevant mechanisms and (ii) how we should ascribe features of a perceptual experience to those mechanisms. Undertaking these two tasks with respect to the purely auditory experience is the subject of the next section.

2.3 The target notion of a purely auditory experience

This, then, is how I propose we understand purely auditory experiences: A purely auditory experience is one arising from the stimulation, by sound waves, of the basilar membrane (or functionally equivalent receptor organ) and which receives no informational input from the other senses or from other non-auditory background representations. The basilar membrane is found in the cochlea (in the inner ear). Sound waves entering the ear are transmitted to the basilar membrane via the eardrum and the tiny bones of the inner ear. The resulting perturbations of the basilar membrane stimulate hair cells arranged along the length of the membrane, which is tonotopically structured—the closer together two regions of the basilar membrane are, the more similar the frequencies they respond to. In this way, the basilar membrane breaks complex incoming sound waves into their constituent frequencies and registers the intensity of those independent constituents (more on this in §3.1.2). Stimulation of these hair cells leads to the release of neural signals transmitted along the auditory nerve to the brainstem, from which they are passed along to the auditory cortex.

As a first pass, we can say that a feature of the proximal stimulus (a sound wave) is encoded just in case stimulation states of the basilar membrane reliably track the presence of that feature in the course of its normal functioning. The stimulation states can be local (at a single hair cell along the basilar membrane) or global (at multiple hair cells along the basilar membrane). Global encoding is governed by algorithms specifying the conditions under which isolated portions of the stimulation of the membrane are to be considered together as potentially encoding some feature of the proximal stimulus. The conditions are specified in terms of the information encoded by each transducer, the spatial relations among transducers on a given receptor organ, and temporal relations among the activation states of transducers. This covers the cases in which locally encoded features are brought together to encode further features of the proximal stimulus. Note, the spatial arrangement of the transducers on the basilar membrane is not, itself, encoded. Subsequent processing utilizes this spatial arrangement without ever explicitly representing it—it is not one of the things that gets processed.

We'll say—again as a first pass—that features of objects and events in the distal environment, and the relations they bear to other objects (including spatial relations), are encoded just in case they are (nominally) determined by encoded features of the proximal stimulus. This guarantees that, whenever we have the determining proximal features (which the receptors evolved to detect), we will have the distal – that is, the presence of these proximal features entails the presence of the distal feature. The question of whether there are distal spatial features that are so determined by the encoded features of the proximal stimulus (sound wave) will be the focus of §3. In particular I focus on the spatial features of vantage point relative distance and direction.

One might object that my reliance on determination in the encoding of distal features is overly stringent. After all, it is common in accounts of mental contents to allow background conditions relevant to the selection or performance of a given perceptual mechanism's function—what Dretske ([1981]) calls 'channel conditions'—to play a role in content determination. I do not rule these out of my account, but I treat them as 'implicit information'—informational assumptions built into the operations a perceptual mechanism performs on its inputs (such as the implementation of one of the algorithms mentioned above). Implicit information is not input to or output of the perceptual mechanism or any of its sub-mechanisms, though it factors in the transformation of input to output. That is, implicit information is not explicitly represented during the performance of those operations the way that encoded information is (Shea [2015]). We can now refine our first pass account of distal feature encoding in this way: Features of objects and events in the distal environment are encoded just in case they are determined by the encoded features of the proximal stimulus in conjunction with the relevant implicit information.

We can then make a similar refinement to the account of proximal feature encoding: Features of the proximal stimulus are encoded just in case stimulation states of the basilar membrane, along with the relevant implicit information (corresponding to 'channel conditions' enabling normal functioning of the basilar membrane), determine the presence of that feature. Treating channel conditions in this way has the benefit of putting them into the same format (information) as explicit information, facilitating their integration with information processing accounts of the mind. It also allows us to pinpoint the location at which implicit information impacts cognitive processing and to identify its sources. This is particularly important for isolating our target notion of the purely auditory experience. It allows us to determine which implicit information is relevant for the production of a purely auditory experience.

To illustrate, interaural discrepancies, in time and level, between the stimulation of the two basilar membranes serve as some of the primary cues to sound (source) direction. These differences are registered in the superior olivary complex (SOC) of the brainstem, which takes direct input from the auditory nerve. The interaural disparities are then transmitted to the inferior colliculus (IC), in the midbrain. It is there that we see the first evidence of selectivity for head-relative left–right directions

(azimuth) (Salminen *et al.* [2018]; Schnupp and King [1997]; Sterbing *et al.* [2003]; Thompson *et al.* [2006]).¹⁹ If the disparities are not, themselves, sufficient for calculating azimuth—and we will see that they are not—we will need to attribute implicit information (for instance, about the distance between the ears) that enables these transformations in IC.

But is this implicit information best thought of as auditory? It is tempting to answer in the affirmative, given IC's position in the auditory processing stream—virtually all signals carried by the auditory nerve that are passed along to the auditory cortex pass through IC. However, these transformations in IC are modulated by learned correlations with other sensory inputs delivered by downward projections to the IC (and SOC, for that matter) from the auditory cortex (Bajo *et al.* [2010]; Bajo and King [2013]; Brainard and Knudsen [1993]; Budinger *et al.* [2006]; Feldman and Knudsen [1997]; Peterson and Schofield [2007]). So IC does not contribute to the purely auditory experience because its outputs are conditions by learned correlations with non-auditory sensory inputs.²⁰

By contrast, the cochlea does not receive downward projections from the auditory cortex. Granted it does receive downward projections from IC, but these are involved in defensive measures to protect against hearing damage from loud sources—stiffening the basilar membrane and tightening the muscles of the inner ear to dampen the mechanical stimulation of dangerously loud sound waves (Guinan [2018]). There is no evidence that IC passes along spatial information, or other learned correlations supplied by auditory cortex, to the cochlea. The same holds for the auditory nerve. So, any implicit information we need to attribute to the cochlea or the auditory nerve—concerning the tonotopic structure of the basilar membrane, for example—can unproblematically contribute to the purely auditory experience.²¹

The general rule is: To count as part of the basis of the purely auditory experience, a mechanism performing transformations on inputs originating with stimulation of the basilar membrane(s) cannot rely on information—implicit or explicit—provided by non-auditory sources (such as learned correlations with other sensory inputs). The account of encoding, above, ensures that the explicit (encoded) information we are considering is exclusively auditory, provided the implicit information implicated in encoding explicit information concerning proximal features is. So, in

¹⁹ It is also there that we see the first evidence for selectivity to sound (source) elevation. This selectivity does not rise to the level of a map of auditory space. There is no topographic arrangement of IC corresponding to azimuth (or elevations), despite the selectivity for each. There is such a map in the superior colliculus (SC)—a multisensory midbrain structure downstream of IC (Sterbing *et al.* [2002], [2003]; King and Palmer [1982], [1983]). Given the multisensory inputs to SC, its contributions must be excluded from the purely auditory experience.

²⁰ Even if innate implicit information is still needed, the status of IC as purely auditory is not as secure as we might have thought because it is impacted by non-auditory sensory input. Recall that there are no generally agreed upon criteria for differentiating the senses (beyond their dedicated peripheral receptors). Given this, it is unclear whether IC counts as auditory or multisensory. Since the goal is to isolate a notion of audition that is uncontroversially auditory, we need to exclude IC (and any innate information we attribute to it).

²¹ See n.27 for instances where we will want to ascribe such information to auditory nerve fibres.

arguing for the aspatialism of the purely auditory experience, we will need to show that there is (a) no non-auditory implicit information is implicated in the encoding of proximal features, and (b) there is no purely auditory implicit information that could combine with encoded features of the proximal stimulus to determine locational features of sounds (sources), thereby giving rise to an experience of sounds as located.

Call the sensory modality associated with the encoding of features of sound waves registered by the basilar membrane(s), in conjunction with the implicit information carried by wholly auditory mechanisms, ‘minimal audition’.²² Minimal audition is minimal insofar as it is the most restricted notion of audition (individuated in terms of receptors and stimulus) that is still recognizable as such.²³ The purely auditory experience is an experience that receives all of its informational input from minimal audition. I will argue for Aspatialism about Minimal Audition:

Aspatialism about Minimal Audition (AMA): Minimal audition cannot encode particular vantage point relative distances or directions of sounds (or their sources) in the distal environment.

With a modest assumption, AMA can be leveraged into a defence of aspatialism about the purely auditory experience. The assumption concerns the relationship between encoded features, representational contents, and phenomenal character: For the purely auditory experience to represent one of its objects (a sound or its source) as having a given feature or to make it seem as though one of its objects has a given feature, minimal audition must be able to encode that feature. And for the experience to be accurate, it must be encoding that feature in fact.²⁴ This follows from the fact that the purely auditory experience only receives informational input from the encoding capacities of minimal audition, which allows us to set aside correlations to spatial information provided from elsewhere (personal or subpersonal, learned or innate, sensory or premodal).

²² This method of defining minimal audition bears affinities with proposals for individuating sensory modalities found in (Keeley [2002]; Matthen [2015]).

²³ Which, I must stress again, is not to say that it is one of the acceptable ways of characterizing our everyday notion of audition. It is to say that minimal audition is recognizable as some form of audition. Advocates of action-constituting accounts of perception might object that, in ruling out those inputs that allow us to track our movements, we cease to have something that is recognizable as auditory. However, action-based views have been developed largely as a way of dealing with the perspectival aspects of visual experience and our ability to abstract away from those aspects. They do not support the claim that visual experience *per se* is impossible in the absence of action but only that perception of the intrinsic spatial features of visual objects is impossible in the absence of action (Evans [1982], [1985]; Grush [2000], [2007]; Noe [2004]; Schellenberg [2007], [2010]). Furthermore, there is no straightforward translation of this into the auditory realm: we don’t hear sounds as having shapes (intrinsic or perspectival) and, even if we could hear the shapes of sources, the standard view is that sounds are the primary objects of audition and act as intermediaries between perceivers and the physical objects and events that are sound sources; but see (Kulvicki [2016]).

²⁴ A feature’s being encoded does not guarantee that it would be a constituent of the representational content of the purely auditory experience, only that it could be. What decides whether encoded distal or proximal features (or both) get represented in the content of the experience will not be addressed here.

To illustrate using a different modality: The retinal stimulation arising from looking at, say, a tennis ball will yield a Marrian $2\frac{1}{2}$ D sketch. That sketch is the basis for presenting the 3D shape of the ball as spherical. But to go from the $2\frac{1}{2}$ D sketch to the 3D shape goes beyond the resources supplied by the retinal stimulation, itself—and perhaps beyond the resources of even our ordinary notion of vision (see n.23). The retinal stimulation underdetermines the 3D shape of the ball. The occluded portion of the ball might be flat, concave, or any number of other shapes. Given this assumption, AMA entails Restricted Auditory Aspatialism:

Restricted Auditory Aspatialism (RAA): A purely auditory experience could not present its objects as being located in some particular direction or at some particular distance, relative to the perceiver.

If RAA is true, the phenomenal objection doesn't apply to the purely auditory experience. However, the action is clearly at the level of what gets encoded by minimal audition. And so the focus will be on establishing AMA.

There is no need to determine whether or not we ever have purely auditory experiences, as I describe them. The point is that, if the specification of the relevant auditory mechanism just is minimal audition and RAA is true, then it couldn't be the case that the experience of sounds as located in a given direction or at a given distance were properly attributed to audition, alone, because no experience whose only informational input is the information encoded in minimal audition can represent its objects as located in this way. This opens up a region of logical space heretofore uncharted in the debate over spatial audition—the region between pure and ordinary auditory experiences (and extending to their vague outer boundaries)—making the new line of defence for unrestricted auditory aspatialism available (§2.2, §4), provided RAA is true. The next section establishes that it is (by way of AMA).

3 Spatial Hearing

Given the way I have characterized minimal audition, we should begin with a consideration of the features of sound waves (the proximal stimulus) that the basilar membrane can encode: intensity, frequency, waveform, onset time and duration. Intensity is, roughly, the amount of energy in the wave.²⁵ Intensity correlates with amplitude²⁶ which, in standard graphical representations, is given by the height of the peak of the wave form above (or below) equilibrium (zero). Frequency is the number of periodic repetitions in the wave in a given time interval. Waveform is the shape of the wave within a period. In graphical representations this will be the recurring (at the wave's frequency) pattern of peaks and valleys of the wave. Duration is the time span from onset to decay of the wave, measured at a stationary point.

²⁵ More precisely, it is the amount of energy per unit area in a direction perpendicular to that area.

²⁶ It is proportional to the square of amplitude.

We can get an intuitive grasp on these features of sound by noting their psychological correlates—the features of ordinary auditory experiences with which they are associated. Intensity is experienced as the loudness of the auditory event, frequency as its pitch, and waveform as its timbre (the ‘sound quality’—that feature of sounds that allows us to distinguish, say, a saxophone from a trumpet producing the same pitch at the same intensity).²⁷ It is important to keep in mind, though, that these psychological correlates are not relevant for determining spatial attributes of sounds. Their underlying bases (features of sound waves) are. Information about these features of sound waves exhausts the sorts of information about the proximal stimulus that the basilar membrane encodes.

Intensity is encoded by the degree of stimulation of the basilar membrane, frequency by the location of the stimulation on the basilar membrane. In the case of spectrum, algorithms consider the locations of the stimulation (as encoding the individual constituent frequencies of the complex stimulus) as well as microfluctuations in intensity and frequency.

Of these features, intensity and waveform vary with distance and direction while frequency and duration do not.²⁸ Differences in onset time and intensity at the two ears vary with direction (and, in limited circumstances, distance). So we should look to intensity, waveform, and these binaural discrepancies for possible sources of spatial information regarding the distal location of sounds or sources in minimal audition. What I will show is that each of these putative cues to distance or direction underdetermines distance/direction in minimal audition. Rather than encoding distance/direction, they pose inverse problems in which any one of an infinite number of pairings of distance/direction with some other feature of the environment, sound, or source would account for that cue. Furthermore, in each case the infinite pairing includes the possibility that the sound is located right

²⁷ The mapping of attributes of sound waves to their psychological correlates is not quite as straightforward as I have represented it. For instance, perceived pitch does not always correspond to the lowest frequency component of a complex sound. In the missing fundamental phenomenon, for example, there is no energy at the frequency corresponding to the perceived pitch—it is ‘supplied’ by the auditory system on the basis of implicit information concerning the distribution of frequencies in complex sounds. Relatedly, perceived pitch can vary from that associated with the lowest sounding frequency when the partials are slightly inharmonic (Terhardt [1979]; Terhardt *et al.* [1982]). This raises a worry: if the relevant mechanisms carrying the implicit information necessary for pitch perception are not part of minimal audition, the perhaps minimal audition will not be recognizably auditory, pitch being (intuitively) a fundamental component of auditory experience. However, the relevant operations are performed in the auditory nerve fibres (Cariani and Delgutte [1996]; Chialvo [2003]). The auditory nerve, like the basilar membrane, does not receive input from other senses. So it can be the site of the mechanisms implementing the frequency-based algorithms for extracting perceived pitch from the stimulation of the basilar membrane. Similarly, the physical characteristics determining timbre are encoded by auditory nerve fibres (Town & Bizley [2013]). Having settled this worry, the simplifications will pose no further complications here.

²⁸ I mean here that these vary with changes in distance to a stationary source. There is, of course, the doppler effect for moving sources or auditors. But most sounds we hear are not both of sufficient duration and produced by sources moving fast enough to produce a significant doppler effect. Furthermore, the doppler effect will be subject to the same sort of worries that I raise for intensity and spectrum based cues to distance—it is confounded by changes in the intrinsic frequency of the source, and the amount of doppler shift relative to distance will depend on the speed and angle of approach or recession (Wightman and Jenison [1995], pp. 385–394). This information could be extracted from auditory experience by combining doppler cues with interaural time differences, but see §3.2 for problems with this approach to encoding the relevant spatial features in minimal audition.

where the perceiver is—at no distance and in no direction. Each of these cases is consistent with the sound being anywhere and so cannot be the sole basis of the localization phenomenology cited in the objection, which depends on sound being experienced as at a distance or in some direction.

I will argue that these inverse problems remain unsolved in minimal audition. If I am right about this, minimal audition fails to encode any particular distance or direction: The encoded features of the proximal stimulus (the putative cues) do not wholly determine vantage point-relative distance or direction, and AMA is true. Hence, given our assumption, RAA is true, too.

3.1 Monaural cues

I begin with a discussion of the monaural cues to direction and distance based in intensity and waveform. Once we get clear on what these cues are, we will see that they are not sufficient to locate sounds in distance or direction, and that they cannot work together to do so. (I argue that they cannot be coordinated with binaural cues to locate sounds in distance or direction without appealing to information gathered from some non-auditory source in §3.2.)

3.1.1 Intensity

In ordinary audition²⁹ the intensity of a sound wave is one obvious cue to the distance of the sound (source). (Intensity shows up in experience as loudness: the Whistler sounds quite faint at my position as I walk about my neighbourhood—much fainter than the whoosh of a passing car. But, given the distance it must travel (it doesn't come from anywhere in my neighbourhood), the Whistler must be much louder at its source than the whoosh is just outside the car.) Call the intensity of a wave at its source the initial intensity. Call the intensity as measured at the listener's position the perspectival intensity.

That perspectival intensity decreases with distance from the sound source follows from the inverse square law—as the spherical wavefront expands, the energy is distributed over a larger and larger area, so less energy is carried at a given point on its surface.³⁰ Therefore, the perspectival intensity of a wave varies with distance from its source to the listener. But it also varies with the intensity of the sound wave measured at its point of origin. And this means that perspectival intensity underdetermines distance. For any given perspectival intensity, there is an infinite set of distance,

²⁹ In this section I will set aside the individuation worries raised in §2.2 and use 'ordinary audition' in its rough-and-ready intuitive sense—the same sense of audition that is invoked in the phenomenal objection. Whether or not this intuitive notion actually corresponds to the best way (or even an acceptable way) of individuating audition will be picked up again in §4.

³⁰ This is an idealization, which assumes the sound source is a point. Of course, sound sources are generally not points, and the geometry of the sound source impacts the rate of decay. However, the idealization only makes things more difficult for me as it prescind from further complications undermining the determination of distance from intensity (for instance, minimal audition doesn't encode the shape of the sound source and so can't determine which equation is relevant for determining distance from the encoded intensity). Similarly for the impact of environmental conditions and the impact of the resistance of the medium.

initial intensity pairs that might correspond to it. But then the initial intensity must be tracked for this cue to suffice for encoding distance (Zahoric *et al.* [2005]).³¹ Minimal audition encodes distance via intensity, alone, only if it encodes both initial and perspectival intensity. But minimal audition doesn't encode initial intensity. The only intensity we get from stimulation of the basilar membrane is perspectival intensity (because the basilar membrane doesn't measure how much it is being stimulated at some location it doesn't occupy). Therefore, minimal audition doesn't encode distance merely in virtue of encoding intensity.

One might worry that auditory implicit information concerning reliable correlations between intensity and distance that could be used to mitigate this indeterminacy. However, the empirical evidence locates the first point at which auditory processing is independently sensitive to distance (that is, the point at which information from distance cues is transformed into information about distance) in the so-called 'where' pathway of the auditory cortex—especially in the planum temporale and the posterior superior temporal gyrus—which responds to location and motion (Kolarik *et al.* [2012], [2016]).³² Not only does this raise the worry that this processing is already conditioned by the multisensory inputs to IC—through which nearly all signals originating from auditory stimulation pass before reaching the auditory cortex—but these areas of the auditory cortex receive direct input from other senses.³³ In fact, they are implicated in cross-modal auditory spatial phenomenon such as the ventriloquism effect (Callan *et al.* [2015]; Zahoric *et al.* [2005]). Given this, they must be excluded from the mechanisms of minimal audition.³⁴ (And any implicit information they carry doesn't factor in minimal auditory processing.)

The result is that minimal audition will encode the infinite disjunction of possible locations of the sound source for the given perspectival intensity. But this is far too imprecise to give rise to the localization phenomenology cited in the phenomenal objection. Even if one thinks that this information will be enough to give rise to an experience of sounds as space occupiers, this phenomenology will be consistent with the sound being located right where the perceiver is because the information on which it is based is consistent with that location. So this phenomenology won't rule out sounds being identified with or instantiated by pressure waves. This gives us a way of

³¹ This is why estimates of distance from intensity are more accurate for familiar sounds. We know, roughly, the intrinsic intensities of familiar sounds.

³² 'What' and 'where' pathways have been proposed for audition (Rauschecker and Tian [2000]; Ahveninen *et al.* [2006]) on the model of the two visual system hypothesis, which associates the ventral visual processing stream with object recognition and the dorsal visual processing stream with vision for action (Goodale and Milner [1992]; Milner and Goodale [2006]).

³³ The view that the traditional sensory cortices are, in fact, multisensory brain regions is increasingly common among neuroscientists—see (Budinger *et al.*, [2006]; Schroeder and Foxe [2005]; Shimojo and Shams [2001]).

³⁴ At least barring evidence that the relevant implicit information is both unconditioned by these multisensory inputs and sufficient to resolve the indeterminacy. Such evidence is unlikely to be forthcoming, given the extent of multisensory interactions in auditory processing.

operationalizing the ‘particular’ vantage point relative distances and directions cited in AMA and RAA: The information regarding vantage point relative distance must be determinate enough to give rise to phenomenology (without informational inputs beyond those available to minimal audition) that can be used to decide between competing views on the ontology of sounds.

Contrast this with looking at the moon—we have no precise information about the distance to the moon, but we still see it as occupying (distal) space. We lack this information because the moon is much too far for proprioceptive input regarding the angle of convergence of the eyes to allow triangulation of the distance to the fixated object (the moon). If we rule out all proprioceptive input from vision, all objects will be seen as at some highly indeterminate distance, as we now see the moon. But, given the laws of optics and the assumption that the stimulation of the two retinas have the same source, this range of distances will rule out the visual object’s being located right where the perceiver is.³⁵ This is a crucial difference from the case of intensity (and all the other cues discussed below)—one that those pressing the phenomenal objection need to reconcile if the objection is going to rule out wave views of sounds. The point here is that this reconciliation is not possible for minimal audition.

3.1.2 Spectrum/waveform

Most sound waves are composed of many frequency components (called partials) that are integrated into a single percept according to a set of heuristics (Bregman [1990]).³⁶ The spectrum of a wave is the distribution of energy among its partials.³⁷ The waveform is determined by its spectrum: The individual amplitudes of each partial are combined additively, yielding the peaks and valleys of the waveform. (Spectrum is the main physical correlate of timbre.) Sounds perceived to have a discernible pitch have (relatively) harmonic relations between their partials—that is, the frequency of each partial is an integer multiple of the lowest constituent frequency (but see n.27). Sounds that do not exhibit such harmonic relations between their partials are noises. The spectrum of a sound wave changes with distance from its source (Butler *et al.* [1980]; Little *et al.* [1992]). In particular, higher frequencies are more attenuated with increased distance than are lower frequencies. So, at greater distances, we should expect a wave of a given initial spectrum to have lost more energy in its upper

³⁵ If it turns out that this assumption is not carried (as implicit information) by uncontroversially visual neural mechanisms, then what I am describing here is not the visual analog of minimal audition. The example is merely illustrative, though, and does not depend on whether what I have described is minimal vision or something between minimal and ordinary vision.

³⁶ These heuristics are akin to the algorithms of §2.3. While some of these heuristics will be attributable to minimal audition—for example, those involved in spectrum detection—those involved in the production of more complicated auditory representations likely will not, as they are implemented in the auditory cortex (Christison-Lagay *et al.* [2015]).

³⁷ This should alert us to the fact that intensity can be understood either locally (the intensity of one frequency component of the complex stimulus) or globally (the overall intensity of all the frequency components as captured by one of our algorithms). The discussion of the preceding section refers to the latter.

partials than in its lower partials, allowing spectrum to function as a distance cue in ordinary audition.

Stimulation of the basilar membrane will fail to encode initial spectrum for reasons analogous to those for which it fails to encode initial intensity. Where before we were concerned with the overall level of excitation of the basilar membrane, now we are concerned with the distribution of that excitation in different regions of the basilar membrane. Otherwise the reasoning is the same: Variations in perspectival spectrum are confounded with the initial spectrum of the sound wave at its source. But the basilar membrane encodes the spectrum at the position of the hearer, not at the position of the source. So again we have an infinite set of pairs of distances (including no distance at all) and some feature of the wave—in this case, initial spectra. Spectrum underdetermines distance in minimal audition. And again, implicit information concerning reliable correlations—for instance that, in our environment, sounds with more energy in their lower partials are much more likely to be distant—won't be a help here. As we've already seen in the discussion of distance from intensity, the first evidence of distance sensitivity from auditory cues is found in the the auditory 'where' pathway, which receives input from the other senses and so must be omitted from minimal audition.

Spectrum is also used, in ordinary audition, to identify sound direction in the median plane—the plane perpendicular to and bisecting the interaural axis. This is accomplished by identifying prominent frequency bands in the overall spectrum of the wave. That the particular frequency bands are associated with particular directions is due to the fact that the pinna (the part of the ear sticking out from the sides of the head) and inner ear amplify different frequency ranges in a sound wave depending on its angle of arrival due to their shape (Blauert [1997]).

However, these cues are also confounded by initial spectrum, which might just happen to have prominent frequencies that are associated with a given direction in the median plane. Of course, we could get direction out if we simply had a mapping of prominent frequency bands to directions, but then spectral cues, alone, wouldn't determine direction—spectral cues plus the mapping would. But, such mappings aren't part of minimal audition: Research suggests that spectral cues for direction are mapped to locations in the inferior colliculus (Grothe *et al.* [2010]). But—as we've seen—these mappings are modulated by non-auditory information, and so must be ruled out of the mechanisms of minimal audition.³⁸ Similarly, we could remove the ambiguity if we could track head position through time, observing the changes in the perspectival spectrum as we do so. But that also requires going beyond minimal audition to include proprioceptive input.

3.1.3 Direct to reverberant ratio

³⁸ That spectral cues are modulated by input from the other senses is also supported by research demonstrating that participants outfit with artificial pinnae were initially impaired in their ability to locate sounds in elevation, but were able to correctly localize sounds after a couple days spent wearing the prostheses while going about their normal daily routines (Hoffman *et al.* [1998]; Trapeau and Schönwiesner [2018]).

The direct to reverberant ratio (DRR)—the intensity of the direct signal relative to the intensity of reflections (subsequent stimulation caused by the original sound wave bouncing off a surfaces back to the ear)—has also been shown to be an effective cue to distance in enclosed spaces in ordinary audition (Bronkhorst and Houtgast [1999]; Kolarik *et al.* [2013]; Kopco and Shinn-Cunningham [2011]). DRR distance estimates exploit a feature of audition known as the precedence effect. The initiation of an auditory event is accompanied by a brief inhibitory period (~40 msec) during which further auditory events are not formed.³⁹ Rather, stimulations of the basilar membrane during this period are parsed as providing information about the already initiated auditory event(s). The precedent effect can be captured (in its relevant details) by one of our algorithms: Stimulation of like regions of the basilar membrane occurring within ~40 msec are considered together as encoding some feature of the proximal stimulus (namely DRR).⁴⁰

We can get a grasp on the phenomenon by considering the difference between reverberation and echoes. An echo is experienced as a distinct auditory event. Reverb is not. Both are the result of sound waves bouncing off a surface before reaching our ears. The only difference is the time it takes for them to do this. DRR is a comparison of the intensity of the direct stimulation (which initiates the precedence effect) and the intensity of subsequent reverberant signals, arriving within the scope of the precedence effect. DRR varies with the distance of the sound source from the perceiver and the size and shape of the space in which the sound is heard. In a given space, the DRR will be lower with greater distance (there will be a weaker direct signal relative to the reverberant signal). But the same DRR in a different space will be correlated with a different distance. Yet again we have an infinite list of pairs (now distance and room dimensions) corresponding to our putative distance cue (DRR). DRR underdetermines distance in minimal audition.⁴¹

This underdetermination could be solved if we could encode the shape and size of an enclosed space in minimal audition. But to do so requires encoding the distances and directions to various surfaces in the environment from stimulation of the basilar membrane. The basilar membrane is stimulated by sound waves, so we would need sound waves emanating from each surface of the room—as they do in a reverberant space. And that compounds the initial problem. Now instead of encoding the distance to one sound source, we need to encode the distance to a tremendous number of (effective) sound sources—points along the surfaces of the enclosed space that reflect sound

³⁹ Unless they are substantially louder than the initiating event.

⁴⁰ The precedence effect relies on implicit information; namely, that the the direct and reverberant signals have the same source. However, the effect first shows up in auditory nerve fibres and, so, can be included—along with this implicit information—unproblematically in minimal audition (Brown *et al.* [2015]; Parham *et al.* [1996], [1998]). The same assumption is at play in the binaural cues discussed in §3.2. There, the relevant implicit information is ascribable to the interaural discrepancy detectors in SOC.

⁴¹ Plausibly, DRR does encode information about the volumetricity of the space in which the sound is heard, given the assumption that the reverberant signal is caused by the same sound source. But this does nothing to locate the sound, itself, which could—for all DRR tells minimal audition—be anywhere, including the perceiver’s location.

waves back to the listener. Complexity, aside, DRR is going to need some other distance cue to define the shape and dimensions of the enclosed space in which it is effective and thereby to allow encoding of the distance to a sound source within that space.

Here it is less likely that there are environmental regularities that could factor, as implicit information, in resolving the indeterminacies. But even if there were, present research locates the transformation of DRR into distance in the ‘where’ pathway of the auditory cortex (Kopco *et al.* [2012], [2020]), which we have already excluded from the workings of minimal audition. Therefore, minimal audition doesn’t encode distance wholly in virtue of DRR.

3.1.4 Monaural cue coordination

What about the possibility that the cues work together to encode distance or direction? We have seen that each cue yields an infinite disjunction of values for the (pairs of) features with which it varies. In order for cue coordination to work, the apparent values for one cue must restrict the number of admissible pairs from the disjunctive list of another. For instance, perspectival spectrum—itsself consistent with an infinite disjunction of distance, initial spectrum pairs—will have to eliminate a large number of the distance, initial intensity pairs given by perspectival intensity. And this requires dependencies between the non-spatial features with which these cues vary—that is, it requires that an auditory event with a given perspectival spectrum could not have been produced by a wave of a given initial intensity at a given distance.

But there are no such dependencies. Intensity is a measure of the overall energy of the wave. Spectrum is a measure of the distribution of that energy across the wave’s constituents. For any initial intensity and distance matching a perspectival intensity, there will be an intrinsic spectrum that would match the perspectival spectrum at that distance. Similar remarks apply to DRR. DRR varies with the dimensions of the enclosed space in which the sound is heard, but there simply is no dependency between the initial intensity or spectrum of a sound wave and the dimensions of the space in which it is heard.

Furthermore, reliable correlations found between these cues and the location of sound (sources) will be of no help. Cue coordinations for distance cues utilizing such implicit information—for example, that sound of a certain intensity and spectrum tend to be far away—are still bound by the empirical results showing no selectivity for distance prior to the auditory ‘where’ pathway discussed above. So that is where we should attribute such implicit information, in which case it is not available to minimal audition.

So much for the monaural cues, unless they can combine with the binaural cues to give distance or direction in minimal audition. I will argue below that they can’t. But first I will argue that the binaural cues, themselves, fail to encode distance and direction in the absence of correlations with information not encoded by the basilar membrane.

3.2 Binaural cues

Binaural cues—those depending on the coordination of inputs at two ears (more to the point: at two basilar membranes)—can be admitted to minimal audition by one of our algorithms (akin to the precedence effect, now applied to stimulation across basilar membranes rather than at one basilar membrane).

3.2.1. Interaural level difference and interaural time difference

The binaural cues are interaural level difference (ILD)—the difference in intensity of a sound wave at the right and left ears, due primarily to the acoustic shadow cast by the head at the farther ear—and interaural time difference (ITD)—the difference in arrival time at the right and left ears. They are used to determine the direction of sounds (or sources) in ordinary audition. When the wave has greater intensity and/or arrives sooner at the left ear, for instance, the auditory event will be experienced as being on the left (Brungart *et al.* [1999]; Duda and Martens [1998]; Konishi [1993]). For sources more than a meter away, directional information from ILDs and ITDs generally coincides, so for the present I will proceed as though there is a single binaural cue.⁴² I address the cases where ILDs and ITDs differ in §3.2.2.

ILDs and ITDs can be used to specify a cone-shaped region of space—the so-called ‘cone of confusion’—extending outward from the perceiver’s head directly to the left or right and with its vertex at a point on the interaural axis (see Figure 1, left). ILDs and ITDs vary with the angle of the sound source relative to the interaural axis.⁴³ (Rotating this angle about the interaural axis generates the cone.) Idealizing a bit, a sound wave emanating from anywhere on the surface of a given cone will result in the same ILD/ITD. (The localization of the sound actually falls along the surface of a volume generated by rotating a parabola about the interaural axis. This volume mostly coincides with the cone but intersects the interaural axis between the ears at a point nearer the ear closer to the sound source, whereas the vertex of the cone proper, is at the midpoint of the interaural axis. For any point

⁴² Readers familiar with the duplex theory of sound localization might wonder about these cases of overlap. In the standard duplex theory, ILDs dominate binaural sound localization for sounds above about 1500 Hz while ITDs dominate below that threshold. This is a matter of dominance (which cues are weighted more strongly) not exclusive domains. However, there remains a substantial overlap of effective ranges of ILD and ITD in the traditional duplex views. Furthermore, recent work has helped specify the role of ILD detection in auditory localization throughout the range of hearing, particularly in conjunction with ITDs at frequencies below the traditional duplex theory threshold (Grantham [1984]; Harmann *et al.* [2016]; Yost [1988]). Similarly, sensitivity to, and a role in spatial processing for, ITDs in high frequency sounds above the traditional duplex theory threshold has been demonstrated (Henning [1974]). Note also that most research supporting the duplex theory relied on sine tones—simple tones without partials—to test the effective ranges of ILDs and ITDs, but naturally occurring sounds are complex with many partials at different frequencies that straddle the traditional duplex theory boundary.

⁴³ For a spherical head, $ITD = d/(2c)(\theta + \sin\theta)$, where d is the interaural distance, c is the speed of sound, and θ is the angle of the source relative to the midpoint of the interaural axis. ILDs are more complicated, but a little picture thinking should suffice to show that, for a spherical head, the shadowing effects will vary with angle of arrival.

on the surface of the parabola, the difference in distance from that point to the left ear and from that point to the right ear is constant.)

Even if minimal audition could specify a cone of confusion to produce sound localization phenomenology on the basis of encoded ILD/ITD information (and just this basis), the resulting phenomenology wouldn't place the sound in a particular direction. Each cone has a vertex on the interaural axis—right where the perceiver is, rather than off in some direction. Furthermore, each cone is symmetrical about the interaural axis. ILD/ITDs won't place a sound, say, to the left and somewhat in front of the perceiver. They would place the sound to the left and either somewhat in front of or somewhat behind (or above, or below) the perceiver.

However, minimal audition can't determine a single cone of confusion from ILD/ITD. Consider ITDs, which are determined, not just by the angle of the wave's arrival and the speed of sound, but also by the interaural distance (see n.43). If we are going to solve for the angle of the sound source relative to the interaural axis (and hence for which cone of confusion the ITD corresponds to), then we will need information regarding the interaural distance as well as ITD.⁴⁴ And interaural distance is not available to minimal audition: Recall that the transformation of ITD (registered in the medial superior olive in the brainstem) into azimuth of the sound source occurs in the inferior colliculus. The inferior colliculus receives input from the auditory cortex which carries information about learned correlations between auditory stimulations and other sensory inputs that modulate the transformation of ITD into azimuth. These correlations—between proprioceptive input concerning head and eye movements and the resulting changes in visual and auditory stimulation—will be the basis for any implicit information concerning interaural distance.⁴⁵ As such, we cannot include this information in the operations of minimal audition.

Similarly, to determine the size of the acoustic shadow cast by a head and thereby determine a cone of confusion from ILD (registered in the lateral superior olive), we need to know the dimensions of the head (including the interaural distance). Yet minimal audition does not have access to information concerning the dimensions of the head. These, too, cannot factor into auditory processing prior to the inferior colliculus, where (again) the transformation from interaural disparity to external direction (azimuth) is made.⁴⁶

⁴⁴ A little more picture thinking to illustrate: The parabola specifying the region in which the sound is located will be wider when the interaural axis is longer because we will need to compensate for the increased distance to the far ear by bringing the sound source closer to it to maintain the difference in the distance from the source to the left and right ears. But this changes the egocentric direction to the sound (source).

⁴⁵ After all, children need to be able to update the interaural distance as they grow. So even if they begin with a default representation of interaural distance in IC, it will quickly be superseded by a refined representation that is conditioned by multisensory input.

⁴⁶ These points get further support from studies of adaptation to modified interaural cues in cases of unilateral deafness or hearing impairment and experimental interventions (Javer and Schwartz [1995]; Trapeau and Schönwiesner [2015]).

At best, then, ILD/ITD determines, an infinite range of cones opening to one side of the head or the other in minimal audition. But even this credits minimal audition with too much: minimal audition can't determine if a given ear is the right or left ear because it has no access to information by which the interaural axis can be oriented relative to external space—or even the perceiver's bodily space. (This orientation is one more piece of information necessary to extract spatial information from the putatively auditory cues that will need to be accounted for in the transformations of interaural disparities to azimuth in IC.) So it can't determine whether a given cone—or an infinite range of cones—opens to the left or to the right. Therefore, minimal audition doesn't encode any particular direction from a given ILD/ITD, and ILD/ITD pose no threat to AMA. Every ILD/ITD is consistent with the sound (source) being located anywhere, including right where the perceiver is.⁴⁷

3.2.2 Binaural and monaural cue coordination

At close distances (approximately 1 meter) the directional information extracted, in ordinary auditory experience, by ITDs and ILDs come apart. In these cases, ILDs—in conjunction with directional information determined by ITDs—can determine distance. In particular, ITDs generate a cone of confusion and the discrepant ILD is used to determine a spherical region of space with its centre in line with the interaural axis and its surface intersecting the interaural axis between the ears (nearer the near ear).⁴⁸ The sound source will be located somewhere within this sphere, excepting a smaller spherical region also centred on the interaural axis but shifted towards the ear. (This inner sphere is the sphere that would be generated by an ILD one just noticeable difference below the registered ILD.) The resulting volume is a spherical shell that is thickest directly opposite the ear, thinnest where it intersects the interaural axis between the ears. When superimposed on the cone of confusion, this sphere isolates a torus-shaped section of the cone on which the sound (source) is located (Shinn-Cunningham *et al.* [2000]) (see Figure 1).

Note that this spherical region, is consistent with a sound source at the perceiver's location. But once the spherical region and cone are superimposed to specify a torus-shaped region of space at their intersection in which the sound source might be located, this is no longer (necessarily) the case. However, we must once again contend with the fact that ITDs will not, in minimal audition, determine a single cone of confusion. So the region of intersection is far less precise than it is in ordinary audition. Indeed, ITDs won't rule out any part of the sphere—there will be some cone intersecting a given portion of the sphere for the given ITD. And the calculation for determining this

⁴⁷ For ILD/ITDs of zero, we might think that we can get some directional information; namely, that the sound is located on a plane perpendicular to and bisecting the interaural axis. However, an ILD/ITD of zero is also consistent with sounds surrounding the listener and sounds emanating from right where the listener is.

⁴⁸ These ILD variations follow the inverse square law with the result that any location of the source that maintains the given proportional distance of the source to the left and right ears will produce the same ILD (provided we disregard the impact of the material properties of the head—see below). Taken together these locations form the surface of a sphere.

sphere from ILD requires information regarding the interaural distance, too. So, without information about the interaural distance, we also get an infinite range of spheres. When coupled with the difficulty concerning the orientation of the interaural axis, the result is that the region of possible intersections of cones and spheres is unconstrained.

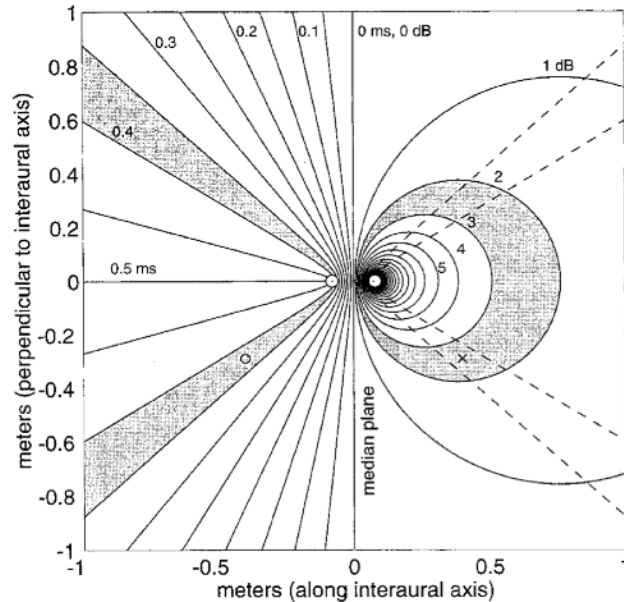


Figure 1. The left half of the diagram represents cones of confusion generated by different ITD values. The shaded areas show a cross-section of the cone that would be generated by a sound source located at 'o'. The right half shows cross-sections of ILD-determined spheres. Dotted lines show the ITD cone for a sound source located at the 'x'. A cross-section of the 'tori of confusion' is found at the intersection of this cone (dotted lines) and the shaded circular region (at 3 dB ILD and .4 ms ITD). (from Shinn-Cunningham et al. [2000], p. 1629)

Furthermore, the material properties of the head—its size, shape, and acoustic absorbency, as well as the position of the ears on its surface—will impact the ILD, and hence, the ILD-ITD discrepancy. And so the same ILD-ITD discrepancy is consistent with different localizations of a sound for differently shaped or constituted heads. In particular, for any ILD-ITD discrepancy, there could be a head so constituted that this discrepancy would result from a sound (source) at the perceiver's location. The result is that minimal audition cannot take advantage of the ILD-ITD discrepancy to determine the distance of a sound (source). So far as minimal audition is concerned, a sound associated with a given ILD-ITD discrepancy could be anywhere, including right where the perceiver is.

This is disanalogous to the stereopsis case (see §3.1.1), in which a visual object was determined to be at an imprecise distance from the perceiver—but at some distance or other—from

retinal disparities along with the assumption that the stimulations had the same source and the laws of optics. In the case of discrepant ILDs and ITDs, we have a disparity and a unity assumption, but there are no laws of nature that take us from there to the fact that the auditory object is some distance from the perceiver. To get that, we also need inputs concerning the interaural distance and material properties of the head. And since minimal audition doesn't encode or implicitly carry this information, it doesn't encode distance (or direction) in virtue of ILD-ITD discrepancies.

What about the coordination of binaural cues with monaural cues? It is unclear what additional information we could get by combining ILD and intensity cues because intensity values recorded at each ear are already factored into the calculation of ILD. There would need to be some spatially relevant information revealed by comparing the two—we would need to be able to restrict the distance or direction of the sound source determined by one by considering the other. For this to work ILD would need to constrain the source of underdetermination for intensity or intensity would need to constrain the source of underdetermination for ILD. But neither manages to do this: the source of underdetermination for intensity is initial intensity. The only constraint a given ILD value places on initial intensity is that the initial intensity be great enough to reach both ears from its point of origin. But there will be an initial intensity that satisfies this condition for a sound source any distance from the perceiver (along the ILD-determined cone of confusion). Similarly, intensity does not constrain the sources of indeterminacy for ILD: it does not limit the interaural distance or material properties of the head.

That leaves DRR and spectrum. Suppose you take DRR measures at both ears and compare them with ILD. ILD is a measure of just the direct signal. If it were not, the spatialization from ILD in an ordinary auditory experience could be confused—the overall measure of energy ($D+R$) could be greater at the farther ear if that ear were much closer to a wall thereby getting greater reverberant energy while the discrepancy in the direct signal remains relatively small. In such a condition, ILD spatialization would place the sound (or source) as in the wrong direction. But then the interaural differences in reverberant intensity vary independently from ILD. They vary with the locations of the sound source, the dimensions of the the space in which it is heard, and the location of the perceiver in that space. ILD varies with the location of the sound source and material properties of the head. The relevant material properties of the head—excepting, to a (too) limited extent, size—are not constrained by, or constraining on, the dimensions of any enclosed space the head happens to occupy. So comparisons of binaural DRR to ILD do not provide any additional information concerning the location of a sound (or source). The same goes for comparisons of binaural DRR to ITD, which varies with location of the sound source and distance between the ears (and which is also not sufficiently constrained by, or constraining on, room dimensions). And, given that DRR doesn't constrain any of the relevant features for either ILD or ITD, it won't help where ILDs and ITDs come apart, either.

Binaural differences in spectrum, since they are law-governed functions of distance, constituent frequency, and the resistance of the medium, will not yield any useful information not already contained in the ILD. The overall level difference will be describable in terms of the level differences between individual frequency components, but that is all. Granted, pinna filtering affects sounds coming from the front differently from those coming from the back, so the spectral direction cue could resolve front–back ambiguity for a given cone of confusion. However, this would require implicit information regarding the relative distributions of innate spectrum types in the environment and the relative front–back effects of pinna filtration, which are not available until IC, at the earliest.⁴⁹ Any implicit information that might factor in the integration of spectral distance cues with the binaural cues—as with all auditory distance cues and their combinations—is not plausibly attributed to auditory processing prior to the ‘where’ pathway. But neither IC nor the ‘where’ pathway are part of minimal audition. And so the coordination of cues will not determine vantage point-relative distance or direction.

Therefore, minimal audition does not encode vantage point-relative distance or direction. AMA is true. And if AMA is true—given our assumption that the content and phenomenal character of a purely auditory experience are dependent on the encoding capacities of minimal audition—RAA is, too: The purely auditory experience cannot present its objects as located at some distance or in some direction in an egocentric frame of reference without drawing on contents provided by some non-auditory source.

4 RAA, unrestricted aspatialism, and the ontology of sound

I now return to the strategy outlined in §2.2 for leveraging RAA into an argument that the phenomenal objection to unrestricted auditory aspatialism is inconclusive, thereby casting doubt on the use of the phenomenal objection to secure results about the ontology of sounds. My goal here is to make the proposal more concrete in light of the details of the putatively auditory spatial cues canvased in §3.

The strategy leans on the fact that there are no generally agreed upon criteria for individuating the senses and the possibility that there are distinctively multisensory phenomenal features of perceptual experiences. These points call into question the association of sound localization phenomenology with audition. It is perfectly plausible that, on the correct (or a correct) individuation of audition, the localization phenomenology is a multisensory feature—only supplied by the operation of audition in conjunction with non-auditory sensory inputs. This is plausible because this phenomenology is not attributable to at least some notion of audition; namely, minimal

⁴⁹ Indeed, there is evidence that such integration of spectral and binaural distance cues is performed in IC (Sterbing *et al.* [2003]). There is also evidence of spectral cues impacting azimuthal calculations in the superior colliculus (SC) of the mouse (Ito *et al.* [2020]). SC is downstream of the IC and receives substantial multisensory input.

audition. So additional mechanisms will need to supplement those of minimal audition in order to produce this phenomenology. But, for each addition, we need to assess whether or not the resulting mechanism is auditory. To do this we need a clear set of criteria for individuating (non-minimal) audition, and we don't have this.

To illustrate, consider again O'Callaghan's ([2017a]) interpretation of Strawson's purely auditory experience—an auditory experience had by someone who has never had experiences in any other sensory modality. (This is a more permissive notion of the purely auditory experience than the one at issue in RAA.) As we've seen, to get the full benefit of the cues discussed in §3, the O'Callaghan-style purely auditory experience will need to be augmented with correlations to sensory inputs in the other senses. For instance, correlations between proprioceptive input regarding head movements and the resulting changes in auditory and visual stimulation could suffice to set the interaural distance and orient the interaural axis with respect to external space, thereby permitting ITD and ILD to determine a single cone of confusion and spectral effects to determine the elevation of a sound (source)—as seems to happen in IC. The question to settle is whether the mechanism resulting from the addition of these learned correlations is best thought of as auditory or multisensory.

On this question, researchers disagree. The multisensory interpretation matches de Vignemont's ([2014], [2018]) etiological account of what it is for a feature to be multisensory.⁵⁰ By contrast, O'Callaghan's ([2015], [2017a], [2017b], [2019]) account of multisensory features requires occurrent stimulation of multiple senses. Provided that the augmented experience doesn't require occurrent stimulation of some other sense, the resulting localization phenomenology will be auditory on O'Callaghan's approach.⁵¹ But auditory aspatialism will be true with respect to the etiological approach, on which the phenomenology cited in the phenomenal objection isn't auditory. Without a clear set of criteria for individuating the (ordinary) auditory mechanism, we can't decide between these approaches. Similar problems arise for any non-sensory inputs that turn out to be involved in extracting distance and direction information from the cues discussed in §3.

The result is that the phenomenal objection is inconclusive. And this inconclusiveness extends to the metaphysical claims that the objection is used to support. The phenomenal objection works in concert with the claim that our perceptual experiences do not systematically deceive us and the—usually unarticulated—assumption that sounds are the objects of audition. If audition presents its objects as located and doesn't systematically deceive us, then its objects must be located. Since those objects are sounds, sounds must be located. But given the competing notions of audition, we must consider the possibility that the objects of these distinct notions might be different, too. For

⁵⁰ This approach has also been adopted by Wong ([2017]) and accords with standard scientific practice.

⁵¹ Briscoe ([2019]) adopts O'Callaghan's approach. See (Wadle [2020]) for a critique of this way of associating features with senses.

instance, the object of an O’Callaghan-style purely auditory experience might be one thing—for instance, a wave—while the object of an experience that adds in learned correlations with other sensory inputs is another—such as an event. If that’s right, the phenomenal objection doesn’t get us to the claim that sounds are events because the relevant phenomenology only appears in an experience whose objects are events, not in an experience whose objects are waves, and we haven’t been given a reason to prefer the former characterization of the auditory experience over the latter.⁵²

To block this line of defence the spatialist will either need to offer an argument that, by ‘sounds’, we mean the objects of the more expansive rather than the less expansive sort of experience or that the objects of both sorts of experience are the same thing, though one gives a more impoverished experience of that thing. Either way, we become embroiled in other contentious debates—over metasemantics and the correct account of mental content, respectively.⁵³ In short, more needs to be said before the phenomenal objection can be taken as supporting ontological claims about the nature of sounds.

5 Conclusion

I have argued that, given the ubiquity of interactions between the (putatively) distinct senses, beginning from the earliest stages of perceptual processing, it is not transparent which features of a perceptual experience ought to be associated with which sensory mechanisms. Next, I introduced a minimal notion of audition, defined so as to guarantee that it would circumvent these complications from cross-modal interactions. I then argued that minimal audition cannot encode vantage point relative distance or direction. I characterized a purely auditory experience as an experience that takes its informational inputs wholly from the information encoded by the minimal auditory mechanism and only relies on unequivocally auditory implicit information in the processing of those inputs. Given that representation of a spatial feature in the purely auditory experience is dependent on the ability of minimal audition to encode that feature, the purely auditory experience can’t present its objects as located in a particular direction or at a particular distance. (Nor does it seem as though sounds are located at a distance in some direction in the purely auditory experience.) So, restricted auditory aspatialism is true.

Given these results, the phenomenal objection to unrestricted auditory aspatialism is inconclusive. We do not presently have an uncontroversial notion of the auditory mechanism to which our experience of sounds as located can be attributed. My experience of the Whistler as

⁵² This bears some affinity with Nudds’s ([2009]) view. However, Nudds allows that both sounds – which are abstract particulars instantiated by sound waves on his view—and sound sources are objects of audition, though only the sources are located. My proposal allows us to avoid the difficult task of explaining why one sort of object of audition is experienced as located but the other is not.

⁵³ This result holds even if we adopt a sensory pluralism (Fulkerson [2014b]; Macpherson [2011b], [2011c], [2014]) that allows both approaches to capture acceptable senses of audition. Each of these acceptable notions may well have distinct objects.

located far away to the Southwest—while very real—might not be due to audition on every reasonable demarcation of the auditory mechanism. For all we can say now, it might not be due to audition on any reasonable demarcation of the auditory mechanisms. But then it is premature to conclude that unrestricted auditory aspatialism is false and to draw further conclusions about the nature of sounds from that falsity. If the debate between spatialists and aspatialists is to be decided by appeal to auditory phenomenology, we will have to shift focus to the demarcation of the auditory mechanism and the sorts of features that the mechanism, on a given demarcation, can capture (and therefore contribute to an experience) in the manner I have pursued with respect to minimal audition.

Acknowledgements

Thanks to Janet Levin, James Van Cleve, John Hawthorne, David Wallace, Irving Biederman, and the anonymous reviewers for this journal.

*School of Philosophy
University of Southern California
Los Angeles, USA
wadle@usc.edu*

References

- Ahveninen, J., Jaaskelainen, I. P., Raij, T., Bonmassar, G., Devore, S., Hamalainen, M., Levanen, S., Lin, F., Sams, M., Shinn-Cunningham, B. G., Witzel, T. and Belliveau, J. W. [2006]: ‘Task-Modulated “What” and “Where” Pathways in Human Auditory Cortex’, *PNAS*, **103(39)**, pp. 14608–14613.
- Bajo, V. M., Nodal, F. R., Moore, D. R. and King, A. J. [2010]: ‘The Descending Corticocollicular Pathway Mediates Learning-Induced Auditory Plasticity’, *Nature Neuroscience*, **13(2)**, pp. 253–262.
- Bajo, V. M. and King, A. J. [2013]: ‘Cortical Modulation of Auditory Processing in the Midbrain’, *Frontiers in Neural Circuits*, **6**, article 114.
- Bayne, T. [2014]: ‘The Multisensory Nature of Perceptual Consciousness’, in D. J. Bennett and C. S. Hill (eds), *Sensory Integration and the Unity of Consciousness*, Cambridge, MA: MIT Press, pp. 15–36.
- Blauert, J. [1997]: *Spatial Hearing* (2nd ed.), Cambridge, MA: MIT Press.
- Brainard, M. S. and Knudsen, E. I. [1993]: ‘Experience-Dependent Plasticity in the Inferior Colliculus: a Site for Visual Calibration of the Neural Representation of Auditory Space in the Barn Owl’, *Journal of Neuroscience*, **13**, pp. 4589–4608.
- Bregman, A. [1990]: *Auditory Scene Analysis: The Perceptual Organization of Sound*, Cambridge, MA: MIT Press.
- Briscoe, R. E. [2008]: ‘Vision, Action, and Make-Perceive’, *Mind and Language* **23(4)**, pp. 457–497.
- Briscoe, R. E. [2016]: ‘Multisensory Processing and Perceptual Consciousness: Part I’, *Philosophy Compass*, **11(2)**, pp. 121–133.
- Briscoe, R. E. [2017]: ‘Multisensory Processing and Perceptual Consciousness: Part II’, *Philosophy Compass*, **12**, pp. 1–13.
- Briscoe, R. E. [2019]: ‘Bodily Awareness and Novel Multisensory Features’, *Synthese*, <https://doi.org/10.1007/s11229-019-02156-2>.
- Bronkhorst, A. W. and Houtgast, T. [1999]: ‘Auditory Distance Perception in Rooms’, *Nature*, **397**, pp. 517–520.
- Brown, A. D., Jones, H. G., Kan, A., Thakkar, T., Stecker, G. C., Goupell, M. J. and Litovsky, R. Y. [2015]: ‘Evidence for a Neural Source of the Precedence Effect in Sound Localization’, *Journal of Neurophysiology*, **114**, pp. 2991–3001.

- Brungart, D. S. and Rabinowitz, W. M. [1999]: ‘Auditory Localization of Nearby Sources: Head-Related Transfer Functions’, *Journal of the Acoustical Society of America*, **106(3)**, pp. 1465–79.
- Brungart, D. S., Drulach, N. I. and Rabinowitz, W. M. [1999]: ‘Auditory Localization of Nearby Sources II: Localization of a Broadband Source’, *Journal of the Acoustical Society of America*, **106(4)**, pp. 1956–68.
- Budinger, E., Heil, P., Hess, A and Sceich, H. [2006]: ‘Multisensory Processing via Early Cortical Stages: Connections of the Primary Auditory Cortical Field with Other Sensory Systems’, *Neuroscience*, **143**, pp. 1065–1083.
- Butler, R. A., Levy, E. T. and Neff, W. D. [1980]: ‘Apparent Distance of Sounds Recorded in Echoic and Anechoic Chambers’, *Journal of Experimental Psychology: Human Perception and Performance*, **6**, pp. 745–750.
- Callan, A., Callan, D. and Ando, H. [2015]: ‘An fMRI Study of the Ventriloquism Effect’, *Cerebral Cortex*, **25**, pp. 4248–4258.
- Calvert, G., Spence, C. and Stein, B. E. [2004]: *The Handbook of Multisensory Processes*. Cambridge, MA: MIT Press.
- Cariani, P. A. and Delgutte, B. [1996]: ‘Neural Correlates of the Pitch of Complex Tones: I. Pitch and Pitch Salience’, *Journal of Neurophysiology*, **76(3)**, pp. 1698–1716.
- Casati, R. and Dokic, J. [2005]: ‘Sounds’, in E. N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, available at <<https://plato.stanford.edu/entries/sounds/>>.
- Casati, R. and Dokic, J. [2009]: ‘Some Varieties of Spatial Hearing’, in M. Nudds and C. O’Callaghan (eds), *Sounds and Perception*, Oxford: Oxford University Press, pp. 97–110.
- Cassam, Q. [2005]: ‘Space and Objective Experience’, in J. L. Bermudez (ed.), *Thought, Reference, and Experience: Themes from the Philosophy of Gareth Evans*, Oxford: Clarendon Press, pp. 258–289.
- Cassam, Q. [2007]: *The Possibility of Knowledge*, Oxford: Oxford University Press.
- Chialvo, D. R. [2003]: ‘How We Hear What Is Not There: A Neural Mechanism for the Missing Fundamental Illusion’, *Chaos*, **13(4)**, pp. 1226–1230.
- Chomanski, B. [2017]: ‘Balint’s Syndrome, Visual Motion perception, and Awareness of Space’, *Erkenntnis*, **83(6)**, pp. 1265–1284.
- Christison-Lagay, K. L., Gifford, A. M. and Cohen, Y. E. [2015]: ‘Neural Correlates of Auditory Scene Analysis and Perception’, *International Journal of Psychophysiology*, **95(2)**, pp. 238–245.
- Coady, C. A. J. [1974]: ‘The Senses of Martians’, *Philosophical Review*, **83(1)**, pp. 107–125.

- Connolly, K. [2014]: ‘Making Sense of Multiple Senses’, in R. Brown (*ed.*), *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience*, New York: Springer, pp. 351–264.
- Dretske, F. [1981]: *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.
- Driver, J. and Noesselt, T. [2008]: ‘Multisensory Interplay Reveals Crossmodal Influences on “Sensory-Specific” Brain Regions, Neural Responses, and Judgments’, *Neuron*, **57**, pp. 11–23.
- Evans, G. [1982]: *Varieties of Reference*. Oxford: Oxford University Press.
- Evans, G. [1985]: *Collected Papers*. Oxford: Oxford University Press.
- Feldman, D. E. and Knudsen, E. I. [1997]: ‘An Anatomical Basis for Visual Calibration of the Auditory Space Map in the Barn Owl’s Midbrain’, *Journal of Neuroscience*, **17(17)**, pp. 6820–6837.
- French, C. [2018]: ‘Balint’s Syndrome, Object Seeing, and Spatial Perception’, *Mind and Language*, **33(3)**, pp. 221–241.
- Fulkerson, M. [2014a]: ‘Explaining Multisensory Experience’, in R. Brown (*ed.*), *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience*, New York: Springer, pp. 365–373.
- Fulkerson, M. [2014b]: ‘Rethinking the Senses and Their Interactions: The Case for Sensory Pluralism’, *Frontiers in Psychology*, **5**, pp. 1–14.
- Ghazanfar, A. and Schroeder, C. E. [2006]: ‘Is Neocortex Essentially Multisensory?’, *Trends in Cognitive Sciences*, **10**, pp. 278–285.
- Goodale, M. H. and Milner, A. D. [1992]: ‘Separate Visual Pathways for Perception and Action’, *Trends in Neuroscience*, **15**, pp. 20–25.
- Grantham, D. W. [1984]: ‘Interaural Intensity Discrimination: Insensitivity at 1000 Hz’, *Journal of the Acoustical Society of America*, **75**, pp. 1191–1194.
- Gray, R. [2013]: ‘Is There a Space of Sensory Modalities?’, *Erkenntnis*, **78(6)**, pp. 1259–1273.
- Grice, H. P. [1962]: ‘Some Remarks about the Senses’, in R. J. Butler (*ed.*), *Analytic Philosophy, First Series*, Oxford: Oxford University Press, pp. 133–153.
- Grothe, B., Pecka, M. and McAlpine, D. [2010]: ‘Mechanisms of Sound Localization in Mammals’, *Physiological Reviews*, **90**, pp. 983–1012.
- Grush, R. [2000]: ‘Self, World and Space: The Meaning and Mechanisms of Ego- and Allocentric Spatial Representation’, *Brain and Mind*, **1(1)**, pp. 59–92.
- Grush, R. [2007]: ‘Skill Theory v.2.0: Dispositions, Emulation, and Spatial Perception’, *Synthese*, **159(3)**, pp. 389–416.

- Guinan, J. J. [2018]: ‘Olivocochlear Efferents: Their Action, Effects, Measurement and Uses, and the Impact on the New Conception of Cochlear Mechanical Responses’, *Hearing Research*, **362**, pp. 38–47.
- Hartmann, W. M., Rakerd, B., Crawford, Z. D. and Zhang, P. X. [2016]: ‘Transaural Experiments and a Revised Duplex Theory for the Localization of Low-Frequency Tones’, *Journal of the Acoustical Society of America*, **139(2)**, pp. 968–985.
- Heil, J. [1983]: *Perception and Cognition*, Berkeley, CA: University of California Press.
- Henning, G. B. [1974]: ‘Detectability of Interaural Delay in High-Frequency Complex Waveforms’, *Journal of the Acoustical Society of America*, **55**, pp. 84–90.
- Hoffman, P. N., Van Riswick, J. G. A. and Van Opstal, A. J. [1998]: ‘Relearning Sound Localization with New Ears’, *Nature Neuroscience*, **1(5)**, pp. 417–421.
- Ito, S., Si, Y., Feldheim, D. A. and Litke, A. M. [2020]: ‘Spectral Cues are Necessary to Encode Azimuthal Auditory Space in the Mouse Superior Colliculus’, *Nature Communications*, **11**, article 1087.
- Javer, A. R. and Schwartz, D. W. [1995]: ‘Plasticity in Human Directional Hearing’, *Journal of Otolaryngology*, **24**, p. 111.
- Kant, I. [1768 (1991)]: ‘On the First Ground of the Distinction of Regions of Space’, in J. Van Cleve and R. E. Frederick (eds), *The Philosophy of Left and Right*, Boston, MA: Kluwer Academic Publishers, pp. 27–33.
- Keeley, B., [2002]: ‘Making Sense of the Senses: Individuating Modalities in Humans and Other Animals’, *Journal of Philosophy*, **99(1)**, pp. 5–28.
- King, A. J. and Palmer, A. R. [1983]: ‘Cells Responsive to Free-Field Auditory Stimuli in the Guinea Pig Superior Colliculus: Distribution and Response Properties’, *Journal of Physiology*, **342**, pp. 361–381.
- Kolarik, A. J., Cirstea, S. and Pardhan, S. [2013]: ‘Discrimination of Virtual Auditory Distance Using Level and Direct-to-Reverberant Ratio Cues’, *The Journal of the Acoustical Society of America*, **134**, pp. 3395–3398.
- Kolarik, A. J., Moore, B. C. J., Zahorik, P., Cirstea, S. and Pardhan, S. [2016]: ‘Auditory Distance Perception in Humans: A Review of Cues, Development, Neuronal Bases, and Effects of Sensory Loss’, *Attention, Perception, and Psychophysics*, **78**, pp. 373–395.
- Konishi, M. [1993]: ‘Listening with Two Ears’, *Scientific American*, **268(4)**, pp. 66–73.
- Kopco, N. and Shinn-Cunningham, B. G. [2011]: ‘Effect of Stimulus Spectrum on Distance Perception for Nearby Sources’, *The Journal of the Acoustical Society of America*, **130**, pp. 1530–1541.

- Kopco, N., Huag, S., Belliveau, J. W., Raij, T., Tengshe, C. and Ahveninen, J. [2012]: ‘Neuronal Representations of Distance in Human Auditory Cortex’, *PNAS*, **109(27)**, pp. 11019–11024.
- Kopco, N., Doreswamy, K. K., Huang, S., Rossi, S. and Ahveninen, J. [2020]: ‘Cortical Auditory Distance Representation Based on Direct-to-Reverberant Energy Ratio. *NeuroImage*, **208**, article 116426.
- Kulvicki, J. [2016]: ‘Auditory Perspectives’, in B. Nanay (ed.), *Current Controversies in Philosophy of Perception*, Abingdon: Routledge, pp. 83–94.
- Little, A. D., Mershon, D. H. and Cox, P. H. [1992]: ‘Spectral Content as a Cue to Perceived Auditory Distance’, *Perception*, **21**, pp. 405–416.
- Maclachlan, D. L. C. [1989]: *Philosophy of Perception*. Englewood Cliffs, NJ: Prentice Hall.
- Macpherson, F. [2011a]: ‘Cross-Modal Experiences’, *Proceedings of the Aristotelian Society*, **111(3)**, pp. 429–468.
- Macpherson, F. [2011b]: ‘Individuating the Senses’, in F. Macpherson (ed.), *The Senses: Classical and Contemporary Readings*, Oxford: Oxford University Press, pp. 3–36.
- Macpherson, F. [2011c]: ‘Taxonomizing the Senses’, *Philosophical Studies*, **153(1)**, pp. 123–142.
- Macpherson, F. [2014]: ‘The Space of Sensory Modalities’, in D. Stokes, M. Matthen, and S. Biggs (eds), *Perception and Its Modalities*, Oxford: Oxford University Press, pp. 434–463.
- Malpas, R. M. P. [1965]: ‘The Location of Sound’, in R. J. Butler (ed.), *Analytical Philosophy* (2nd series), Oxford: Basil Blackwell, pp. 131–144.
- Mandrigin, A. [2021]: ‘Multisensory Integration and Sense Modalism’, *The British Journal for the Philosophy of Science*, **72(1)**, pp. 27–49.
- Matthen, M. [2010]: ‘On the Diversity of Auditory Objects’, *Review of Philosophy and Psychology*, **1(1)**, pp. 63–89.
- Matthen, M. [2014]: ‘Active Perception and the Representation of Space’, in D. Stokes, M. Matthen, and S. Biggs (eds), *Perception and Its Modalities*, Oxford: Oxford University Press, pp. 44–72.
- Matthen, M. [2015]: ‘Individuation of the Senses’, in M. Matthen (ed.), *Oxford Handbook of the Philosophy of Perception*, Oxford: Oxford University Press, pp. 567–586.
- Milner, A. D. and Goodale, M. H. [2006]: *The Visual Brain in Action* (2nd ed.). Oxford: Oxford University Press.
- Nelkin, N. [1990]: ‘Categorising the Senses’, *Mind and Language*, **5**, pp. 149–165.
- Noe, A. [2004]: *Action in Perception*, Cambridge, MA: MIT Press.
- Nudds, M. [2001]: ‘Experiencing the Production of Sounds’, *European Journal of Philosophy*, **9**, pp. 210–229.

- Nudds, M. [2004]: ‘The Significance of the Senses’, *Proceedings of the Aristotelian Society*, **104(1)**, pp. 31–51.
- Nudds, M. [2009]: ‘Sounds and Space’, in M. Nudds and C. O’Callaghan (eds), *Sounds and Perception*, Oxford: Oxford University Press, pp. 69–96.
- Nudds, N. [2010]: ‘What are Auditory Objects?’, *Review of Philosophical Psychology*, **1**, pp. 105–122.
- O’Callaghan, C. [2007]: *Sounds*. Oxford: Oxford University Press.
- O’Callaghan, C. [2008]: ‘Seeing What You Hear: Cross-Modal Illusions and Perception’, *Philosophical Issues*, **18**, pp. 316–338.
- O’Callaghan, C. [2010]: ‘Perceiving the Locations of Sounds’, *Review of Philosophical Psychology*, **1**, pp. 123–140.
- O’Callaghan, C. [2012]: ‘Perception and Multimodality’, in E. Margolis, R. Samuels, and S. Stich (eds), *Oxford Handbook of Philosophy of Cognitive Science*, Oxford: Oxford University Press, pp. 92–117.
- O’Callaghan, C. [2014a]: ‘Intermodal Binding Awareness’, in D. J. Bennett and C. S. Hill (eds), *Sensory Integration and the Unity of Consciousness*, Cambridge, MA: MIT Press, pp. 73–103.
- O’Callaghan, C. [2014b]: ‘Not All Perceptual Experience Is Modality Specific’, in D. Stokes, M. Matthen, and S. Biggs (eds), *Perception and Its Modalities*, Oxford: Oxford University Press, pp. 133–165.
- O’Callaghan, C. [2015]: ‘The Multisensory Character of Perception’, *Journal of Philosophy*, **112(10)**, pp. 551–569.
- O’Callaghan, C. [2017a]: *Beyond Vision*. Oxford: Oxford University Press.
- O’Callaghan, C. [2017b]: ‘Grades of Multisensory Awareness’, *Mind and Language*, **32(2)**, pp. 155–181.
- O’Callaghan, C. [2019]: *A Multisensory Philosophy of Perception*. Oxford: Oxford University Press.
- O’Shaughnessy, B. [2000]: *Consciousness and the World*. Oxford: Oxford University Press.
- O’Shaughnessy, B. [2009]: ‘The Location of Perceived Sound’, in M. Nudds and C. O’Callaghan (eds), *Sounds and Perception*, Oxford: Oxford University Press, pp. 111–125.
- Orioli, G., Bremner, A. J. and Farroni, T. [2018]: ‘Multisensory Perception of Looming and Receding Objects in Human Newborns’, *Current Biology*, **28(22)**, pp. R1294–R1295.
- Palmer, A. R. and King, A. J. [1982]: ‘The Representation of Auditory Space in the Mammalian Superior Colliculus’, *Nature*, **299**, pp. 248–249.

- Parham, K., Zhao, H. B. and Kim, D. O. [1996]: Responses of Auditory Nerve Fibers of the Unanesthetized Decerebrate Cat to Click Pairs as Simulated Echoes', *Journal of Neurophysiology*, **76**, pp. 17–29.
- Parham, K., Zhao, H. B. and Kim, D. O. [1998]: 'Responses of Anteroventral Cochlear Nucleus Neurons of the Unanesthetized Decerebrate Cat to Click Pairs as Simulated Echoes', *Hearing Research*, **125**, pp. 131–146.
- Pasnau, R. [1999]: 'What Is Sound?', *Philosophical Quarterly*, **49**, pp. 309–324.
- Peterson, D. C. and Schofield, B. R. [2007]: 'Projections from Auditory Cortex Contact Ascending Pathways that Originate in the Superior Olive and Inferior Colliculus', *Hearing Research* **232(1-2)**, pp. 67–77.
- Rauschecker, J. P. and Tian, B. [2000]: 'Mechanisms and Streams for Processing of “What” and “Where” in Auditory Cortex', *PNAS*, **97(22)**, pp. 11800–11806.
- Roxbee-Cox, J. W. [1970]: 'Distinguishing the Senses', *Mind*, **79**, pp. 1530–1550.
- Salminen, N. H., Jones, S. J., Christianson, G. B., Marquardt, T. and McAlpine, D. [2018]: 'A Common Period Representation of Interaural Time Differences in Mammalian Cortex', *Neuroimage*, **167**, pp. 95–103.
- Schellenberg, S. [2007]: 'Action and Self-Location in Perception', *Mind*, **116(463)**, pp. 603–632.
- Schellenberg, S. [2010]: 'Perceptual Experience and the Capacity to Act', in N. Gangopadhyay, M. Madary, and F. Spicer (eds), *Perception, Action, and Consciousness*, Oxford: Oxford University Press, pp. 145–160.
- Schnupp, J. W. H. and King, A. J. [1997]: 'Coding for Auditory Space in the Nucleus of the Brachium of the Inferior Colliculus in the Ferret', *Journal of Neurophysiology*, **78**, pp. 2717–2731.
- Schwenkler, J. [2012]: 'Does Visual Spatial Awareness Require the Visual Awareness of Space?', *Mind and Language*, **27(3)**, pp. 308–329.
- Schroeder, C. E. and Foxe, J. [2005]: 'Multisensory Contributions to Low-Level, “Unisensory” Processing', *Current Opinion in Neurobiology*, **15**, pp. 454–458.
- Shea, N. [2015]: 'Distinguishing Top-Down from Bottom-Up Effects', in D. Stokes, M. Matthen, and S. Biggs (eds), *Perception and Its Modalities*, Oxford: Oxford University Press, pp. 73–91.
- Shimojo, S. and Shams, L. [2001]: 'Sensory Modalities are Not Separate Modalities: Plasticity and Interactions', *Current Opinion in Neurobiology*, **11**, 505–509.
- Shinn-Cunningham, B. G., Santarelli, S. and Kopco, N. [2000]: 'Tori of Confusion: Binaural Localization Cues for Sources within Reach of the Listener', *Journal of the Acoustical Society of America*, **107(3)**, pp. 1627–1636.

- Sorensen, R. [2008]: *Seeing Dark Things: The Philosophy of Shadows*. Oxford: Oxford University Press.
- Spence, C. [2016]: ‘Oral Referral: On the Mislocalization of Odours to the Mouth’, *Food Quality and Preference*, **50**, pp. 117–128.
- Spence, C. and Bayne, T. [2014]: ‘Is Consciousness Multisensory?’, in D. Stokes, M. Matthen, and S. Biggs (eds), *Perception and Its Modalities*, Oxford: Oxford University Press, pp. 95–131.
- Sterbing, S. J., Hartung, K. and Hoffmann, K-P. [2002]: ‘Spatial Tuning in the Superior Colliculus of the Guinea Pig in a Virtual Auditory Environment’, *Experimental Brain Research*, **142**, pp. 570–577.
- Sterbing, S. J., Hartung, K. and Hoffmann, K-P. [2003]: ‘Spatial Tuning to Virtual Sounds in the Inferior Colliculus of the Guinea Pig in a Virtual Auditory Environment’, *Journal of Neurophysiology*, **90**, pp. 2648–2659.
- Strawson, P. F. [1959]: *Individuals*. London: Methuen Press.
- Terhardt, E. [1979]: ‘Calculating Virtual Pitch’, *Hearing Research*, **1**, pp. 155–182.
- Terhardt, E., Stoll, G. and Seewann, M. [1982]: ‘Algorithm for Extraction of Pitch and Pitch Salience from Complex Tonal Signals’, *Journal of the Acoustical Society of America*, **71(3)**, pp. 679–688.
- Thompson, S. K., von Kriegstein, K., Deane-Pratt, A., Marquardt, T., Deichmann, R., Griffiths, T. D. and McAlpine, D. [2006]: ‘Representation of Interaural Time Delay in the Human Auditory Midbrain’, *Nature Neuroscience*, **9**, pp. 1096–1098.
- Town, S. M. & Bizley, J. K. [2013]: ‘Neural and Behavioral Investigations into Timbre Perception’, *Frontiers in Systems Neuroscience*, **7**, article 88.
- Trapeau, R. and Schönwiesner, M. [2015]: ‘Adaptation to Shifted Interaural Time Differences Changes Encoding of Sound Location in Human Auditory Cortex’, *NeuroImage*, **118**, pp. 26–38.
- Trapeau, R. and Schönwiesner, M. [2018]: ‘The Encoding of Sound Source Elevation in the Human Auditory Cortex’, *Journal of Neuroscience*, **38(13)**, pp. 3253–3264.
- de Vignemont, F. [2014]: ‘A Multimodal Conception of Bodily Awareness’, *Mind*, **123(492)**, pp. 989–1020.
- de Vignemont, F. [2018]: *Mind the Body*, Oxford: Oxford University Press.
- Wadle, D. C. [2020]: ‘Sensory Modalities and Novel Features of Perceptual Experience’, *Synthese*, advance online publication available at <<https://doi.org/10.1007/s11229-020-02689-x>>.
- Wightman, F. L. and Jenison, R. [1995]: ‘Auditory Spatial Layout’, in W. Epstein and S. Rogers (eds), *Perception of Space and Motion* (2nd ed.), San Diego: Academic Press, pp. 365–400.
- Wong, H. Y. [2017]: ‘On Proprioception in Action: Multimodality versus Deafferentation’, *Mind and Language*, **32(3)**, pp. 259–282.

Yost, W. A. and Dye, R. H. [1988]: ‘Discrimination of Interaural Differences of Level as a Function of Frequency’, *Journal of the Acoustical Society of America*, **83**, pp. 1846–1851.

Zahoric, P., Brungart, D. S. and Bronkhorst, A. W. [2005]: ‘Auditory Distance Perception in Humans: A Summary of Past and Present Research’, *Acta Acustica united with Acustica*, **91**, pp. 409–420.