Section IV

Perception and Dreams

Proof

10 Hearing, Seeing, and Music in the Middle

Dan Lloyd

Of all the forms of human inquiry, the one that covers most thoroughly that entity we call "the world" is phenomenology. Its target is the entirety of the-world-as-it-seems-to-us, including all sensory appearances, along with the accompanying apprehensions, reflections, and organizing structures within individual experience. The end is a rich description of this apparent world, comprising both the concrete details and abstract frameworks that sift through consciousness during every waking moment (and many sleeping ones). Thus, to divide phenomenology according to sensory modality is already to impose a distinction that is not nearly so sharp in ordinary experience. Some objects and events are visible and not audible, and some are the converse, but much of the world presents itself as potentially visible *and* audible (and tactile and possibly sensuous in other ways as well). We neither see sights nor hear sounds, but rather we see and hear objects and events that present visible and audible features. Only in special conditions will we attend to the materials of purely visual or auditory sensation divorced from the constant reconstruction of their sources.

Nonetheless we can distinguish between features that we experience through one sense or another, and characterize the aspects of the world that each sense can discern. It is especially illuminating to pull hearing free from vision, because to reflect on hearing helps us see how myopic phenomenology (and neurophenomenology) can be in its assumptions from the point of view of vision. Vision even dominates the metaphors of consciousness—three visual metaphors appear in the previous sentence, for example. To understand the contrast and its implications, this chapter will engage in an expanded phenomenology, one stubbornly grounded in the physical world. To contemporary theorists, perception is embedded, embodied, and enactive (Clark 2008, Noë 2004, Varela, Thompson, and Rosch 1991). The three E's will organize this chapter, though with considerable overlap and crosstalk. Embedding directs our attention to the real environment, the ecology of energies available to the senses of sight and hearing. Embodiment is prominent in the contrasts between the sense organs involved. Action organizes everything else, but the mix with hearing greatly modifies the meaning of action in perception. The transit from environment to sense organ to brain is one of cascading constraints, each level shaping the meaning of the next. Cascading constraints are strikingly predictive of the contents of awareness, so the discussion here will merge into the concept of cascading conscious contents. Finally, one uniquely human intervention in the sonic environment is music. A coda to the paper harmonizes the trio of hearing, seeing, and music in the middle.

1. THE WORLD

We begin with a close look and a close listen to the phenomenal worlds of light and sound. A cup of coffee can get us started. Consider it, first, as an object for visual exploration (in the style of Husserl's 1907 lectures (1973/1997); see also Ihde 2012). From a stationary point of view, the invariants available include the cup's unity as a shape, the covering of the shape with colored patches, its segregation as a figure against a ground, its invisible but apprehended back side, the horizon at the visible edge of the object, the apparent distance from our point of view, the spatiality of the environment, and more. With even the slightest bodily motion, a further bundle of dynamical invariants emerge, as the visible properties shift. Their shifting is elaborately coordinated, preserving the integrity of the cup and, at the same time, implicating bodily kinesthetic awareness. For example, as you lean to the left, the facing side of the cup rotates, occluding a slightly different bit of the desk behind. But it remains a stationary perceptual object because the visual alterations are the physical converse of my bodily movements (including saccades). As you move, some visible properties break loose from the object. For example, as you circle the mug, the reflected highlights shift at half the speed of the rotation, and thereby emerge as a sheen, rather than as part of the intrinsic coloring of the object. Those highlights implicate a source of illumination, and as the sparkle on the cup slides along its surface, so also do many other highlights in the field of view. Implicit in the scene, but depending on the visual constitution of the cup, along with the everyday encyclopedia of background knowledge, are many affordances, as defined by J.J. Gibson (Gibson 1979):

> The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill. . . . [An affordance] refers to both the environment and the animal in a way that no existing term does. It implies the complementarity of the animal and the environment. (ibid., 127)

With its affordances, the cup is "ready-to-hand" (Heidegger 1927/2008, §15), available as a vessel for coffee, or as a paperweight, or as a projectile, or even as an example in this essay. Woven through your visual exploration is the constancy and stability of the visual environment. Objects in this environment are continuously available for ongoing inspection. As a result, the interaction described above can be reenacted, reconfirmed, or varied over time. Although many aspects of the object go beyond the visible features, they depend directly on visual information. We sample and resample them with little reflection.

Now close your eyes and embark on a similar examination of the cup by hearing alone. Strikingly, what is so apparent to the eye is nonexistent to the ear. To audition, none of the properties listed above appear or, if so, only in the vaguest and most rudimentary form. (For example, from hearing you might be aware that you are surrounded by an enclosing space.) Listening with determined attention will not change this essential disconnection. If something—a spoon, for example—strikes the cup, then some of its objective properties appear, but for only the briefest interval. From the ping, you can gather the approximate location of the object and some general features of the resonant surrounding space. From background knowledge, you might conclude that metal is striking ceramic, but not much more. It passes too quickly to probe or examine. You may also hear background noises and attend to them if you choose. But like the audible mug, these intermittent energies emerge sporadically from a field of silence or other sounds. For all these sound sources, the only ongoing inspection available is through the immediate recall of the ephemeral stimulus.

This brief sketch suggests that vision paradigmatically affords a world of objects, while audition affords a world of events. This conspicuous difference between the two landscapes is not due to the differences between sensory systems, but to the landscapes themselves. It happens that our planet is bathed in continuous light energy, allowing objects to be examined and re-examined over time. Sounds in our world are far more sporadic, more like a stroboscopic flash, and thus the information packed in a sound must be extracted from a brief stimulus or a succession of brief stimuli. It also happens that light travels in straight lines, which are sharply occluded by opaque edges, while sound can bounce and bend around corners. These ecological conditions are general but nonetheless contingent. It would be possible for the information landscapes to be reversed:

> The auditory world is like the visual world would be if all objects were very, very transparent and glowed in sputters and starts by their own light, as well as reflecting the light of their neighbors. This would be a hard world for the visual system to deal with. (Bregman 1990, 37)

In the other direction, in special circumstances the landscape of sound acquires ecological features of the visual world. In his memoir *Touching the Rock: An Experience of Blindness*, John Hull describes a scene "illuminated" by rain:

> This evening, at about nine o'clock, I was getting ready to leave the house. I opened the front door, and rain was falling. I stood for a few

> minutes, lost in the beauty of it. Rain has a way of bringing out the contours of everything; it throws a coloured blanket over previously invisible things; instead of an intermittent and thus fragmented world, the steadily falling rain creates continuity of acoustic experience. . . . I think that this experience of opening the door on a rainy garden must be similar to that which a sighted person feels when opening the curtains and seeing the world outside. Usually, when I open my front door, there are various broken sounds spread across a nothingness. . . . The rain presents the fullness of an entire situation all at once, not merely remembered, not in anticipation, but actually and now. The rain gives a sense of perspective and of the actual relationships of one part of the world to another. (Hull 1992, 22–24)

Although these initial explorations are phenomenological, this first stage of analysis displays none of the inaccessible interiority that is often assumed to be the hallmark of phenomenology. Instead, we have begun with consideration of the information landscapes available to vision and audition, an 'ecological phenomenology.' There is as yet no mind-body problem in this scenario; instead, the configurations of energy discussed so far are real, entirely unproblematic for science or philosophy. In short, the distinctions between vision and audition described so far are objective differences; 'real phenomenology' is not an oxymoron. But 'subjectivity' is not thereby excluded. A real subjectivity emerges in two ways. First, the two informational landscapes are subsets of the total array of energies available in the scene. For example, the very same coffee cup and desk comprise an ultraviolet landscape (the scene for some birds), a hypersonic landscape (for bats and rats), a microwave landscape, a neutrino landscape, etc. Obviously the visual landscape is picked out by the sensitivities of human vision, and likewise for audition. In this selection, we have considered the capacities of observers, the subjects doing the observing. But this subjectivity merely selects ecological properties that can cause changes at the sense organs in question. The properties themselves carry on in their mundane reality.

The second appearance of subjectivity is embedded in a point of view. To extract and construct the features of the coffee cup, a point of view is assumed. The bundle of light rays and the ripples of compressed air unpack into the worlds of objects and events only when they are compared at a position in space. These points of view can only be occupied by one observer at a time. Still, this is an unproblematic subjectivity: the point of view is a simple location, and it is a consequence of physics that each location can accommodate just one observer at any time. Gibson analyzes this 'subjectivity' as follows:

> If it is assumed that no two observers can be at the same place at the same time, then no two observers ever have the same surroundings. Hence, the environment of each observer is "private," that is, unique.

> This seems to be a philosophical puzzle, but it is a false puzzle. (Gibson 1979, 43)

Over time, observers occupy multiple points of view and can build a model of a shared environment:

> The available paths of locomotion in a medium constitute the set of all possible points of observation. In the course of time, each animal moves through the same paths of its habitat as do other animals of its kind. Although it is true that no two individuals can be at the same place at the same time, any individual can stand in all places, and all individuals can stand in the same place at different times. Insofar as the habitat has a persisting substantial layout, therefore, all its inhabitants have an equal opportunity to explore it. In this sense the environment surrounds all observers in the same way that it surrounds a single observer. (ibid., 43)

Gibson flips the logical order of 'objective' and 'subjective.' In the traditional picture, a stable, public, objective world causes a unique private subjectivity, but in the Gibsonian scheme, the objective is a construction from all the possible and actual subjective views of observers moving about and interacting with their environment. Subjectivity is the real physical ground of objectivity. The real phenomenology of animals is neither exclusively subjective nor objective—these terms no longer have exclusive denotations.[1]

The ecological focus embeds perception in the world of the embodied perceiver. The two distance senses operate in very different environments; this embedding greatly determines what sensory experience could be like, regardless of the nature of the sense organ. But the environment imposes its heaviest constraints in the forms of active perception usefully engaged when one sense or the other is in play. Visual cognition aims at the construction of meaningful images and their interpretation as arrangements of objects in space; throughout this process, the steady stream of light information is assumed. This allows the mobile exploration of objects and scenes. Differing viewpoints yield elaborations of the scene. A return to previous viewpoints reconfirms continuity. Both actions work in the stable environment of continuous illumination. In short, to explore with the eyes is to move.

Audition builds a phenomenal world as well, but in the normal auditory environment a steady information stream is the exception, found in machine hums, the whistling, pattering weather, and deliberate tones (more on this later). The auditory world demands the clever deconstruction of brief surges of complex sound energies via a process that is very fast, yet extremely versatile in its ability to accumulate information, group sound energies, segregate them in separate streams, and apply subtle causal models of sound sources. All of this happens after the pressure wave of sound energy has passed. Physical movement is too late for circumspection of the auditory event. Audition is almost always an afterthought.

Ecological and 'enactive' psychologists stress the physical movement of perceivers in environments, which is surely essential but also shaped by the long shadow of vision (e.g., Noë 2004, O'Regan 2001). In celebrating the sensorimotor, these theorists oppose a classical cognitive science story of fast hierarchical processing of occurrent information by a stationary observer (e.g., Marr 1982). Both conceptions fit themselves to vision. As discussed above, physical movement effectively adumbrates a perceptual world in which ambient information is mostly continuous and stable—the world of light. This steady and widely illuminated scene is assumed in the classical snapshot as well. The sensory processing hierarchy lights up when the stimulus is present, and as long as the stimulus is present. Again, this occurrent continuity is a property of illuminated scenes. Active hearing modifies both conceptions. Audition brings forward a different kind of 'movement,' namely, the reflective engagement of immediate memory and continuous interpretation, both processes folded together. This is internal and draws on narrower bursts of sensation, but is deliberate and exploratory—just as much an action as bodily movement. Classical vision is the passive reception of an ongoing, occurrent, and simultaneous set of interpretations. Audition lacks that lingering luxury, and so auditory computations are always retrospective and unconstrained by the occurrent stimulus. These computations cut loose from the stimulus and its compulsions. They're more discretionary, deliberate, and contingent, more like action. One might think of perceptual exploration for hearing as movement internalized. It is active rather than passive, as the enactivists propose, but internal, a modification of a cognitivist scheme.

In short, to pre-reflective vision, the standing question is, "What *is* that?" To audition, the eternal question is, "What *was* that?" Audition invokes temporality as a comprehensive human perceptual capacity. In phenomenology, the *locus classicus* of this emphasis is, of course, Husserl (Husserl 1928/1964). The familiar Husserlian model unpacks every moment of consciousness into a temporal act with three aspects. Retention is the immediate recall of the just-happened. Protention is the ongoing anticipation of the about-to-happen. 'Primal impression' names the immediate percepts of the instantaneous Now. All perception is temporal (and all objects temporally extended), but the sputtering ecology of sound makes the necessity of temporality conspicuous. Hearing must look back to sound that no longer exists (retention) and forward to sounds anticipated (protention). It's not surprising that Husserl's prime example of a temporal object is auditory, a melody (Husserl 1928/1964). A melody is only a unity across time, and, to be perceived as a temporal unity, its temporal parts must be bound together in a single conscious percept while time passes. In constituting the Marseillaise as a single experienced entity, Husserlian retention and protention are inescapable. The immensity of the phenomenological task of characterizing temporality is apparent in Husserl's repeated and never satisfactory excursions into temporality, and the multiple adumbrations of temporal phenomenology across the phenomenological tradition (Mensch 2014). All intentional objects are temporal, coffee cups included, yet the valorization of the visual has enabled temporality to be overlooked. Hearing reminds us of the temporal structure of consciousness overall. As Don Ihde emphasizes (2007, 102): "*Sound reveals time*." (See also Mensch 2014, Clarke 2011).

2. WORLD MEETS BRAIN

So far, this reconstruction of phenomenology has traded exclusively in arrays of energy available in typical terrestrial environments. Let us now lift the curtain a little and consider the leading edge of the sensory systems themselves. That frontline is the array of transducers within the sense organs. In their most general function, they collapse distinctions between the senses, since whatever form stimulus energy takes, it is transformed into the lingua franca of axonal signals. But at this point, distinctions emerge due to the systems themselves. The two 'distance senses' are fundamentally different in an invariant phenomenal property that is so pervasive that it may escape notice. Physiologically, both senses are sensitive to a mix of energy at different frequencies, which to one sense appears as color, and to the other as pitch. But when multiple frequencies emanate from a single source (of light or sound), the two senses deliver very different sensations. Two or more frequencies of light will blend into an intermediate perceived color, while multiple frequencies of sound will maintain their differences, resulting in a heard harmony of multiple pitches. If hearing worked like seeing, a complex sound of 262 and 392 Hz (C4 and G4) should be heard as a single tone at 327 Hz (if intensities are matched); entire symphonies would sound like a single wavering tone. If sight worked like hearing, the full spectrum of apparent colors arising from mixtures of the three primary frequencies used in color printing or video screens would disintegrate. The orange in the magazine ad would look like an overlay of pure yellow and pure red, and not at all like the pure spectral orange of the fruit in the market.

The blending of colors is the price we pay for sharp visual acuity, since the four coarsely tuned receptors can be densely packed in the retina, exploiting the sharp geometry of rays of light in order to get color and brightness information at thousands of locations at once. Sound does not propagate in straight lines only. Accordingly, whatever spatial acuity we achieve in the sound world rests on subtle differences in timing and intensity at our two ears. By virtue of the resonant shape of the cochlea, receptors in the basilar membrane are functionally tuned to a continuum of distinct frequencies. To hearing, precise frequency analysis is assigned to approximate locations, while to sight, approximate (mean) frequencies are assigned to precise locations. Swapping these two schemes of proximal sensation would undermine the acuity of perceptual events (the specialty of hearing) and destroy the acuity of objects (the stuff of sight).

Ecological phenomenology has now been augmented with 'transducer phenomenology,' and another aspect of our subjective world has been embodied. Eyes are optimized to collect light information from scenes in which such information is relatively continuous and stable. Ears are optimized to collect as much sound information as possible during the brief duration of the sound itself. In the ascent from the world to the transduced sensory signal, we observe an example of 'cascading constraints.' Ecological conditions and sensory processes combine to limit the information available for perception. These are objective facts. The environment really is a mix of energies at various frequencies, and the sense organs are mere reporters of what exists in their purview. But the slice of reality disclosed is radically shaped by what sense receptors do, and where and how they do it.

A visiting Martian could observe the environmental energies and their transduction as a complex of physical causes and effects. We humans add something crucial—we *experience* the world the Martian describes. The differences between the two sensory processes, even at this first layer, precisely map the fundamental distinctions in the phenomenology of seeing versus hearing. Cascading constraints collapse into hybrid entities; these entities are at the same time the contents of consciousness. Once again, the process is subjective and objective at once. The environment affords a world of illuminated objects and sonic events. The radical differences between the two sense organs reinforce the object/event distinction. The distinction, however, cannot be assigned to any level or stage of processing. Rather, it is emergent through the interaction of both world and transducers. Cascading constraints constitute a single subjective/objective world.

Similar observations follow from the next stage of visual processing, as the receptors for colors and brightness feed into 'opponent process' cells. In this process, the signals of brightness and the long, short, and medium wavelength sensitivities of the cones are remixed along three opponent axes: blue/yellow, red/green, and light/dark (white/black) (Hurvich 1981). Paul Churchland has pointed out the comprehensive match between the three-dimensional space of opponent cell outputs and the parallel space of phenomenal colors (Churchland 2005). Thus, once again, a 'primitive' neural distinction is precisely preserved in 'high-level' conscious awareness. Churchland goes one better, however, using opponent processing and cell fatigue to predict possible new positions in opponent processing space and then devising a method for pushing the visual system into those positions. Phenomenally, this creates novel 'chimerical colors' that can defy description, but are perfectly apparent in conscious experience. Opponent cells do modify the 'raw' outputs of the receptors, however, by compensating for differences in ambient illumination and thereby stabilizing constant colors in perception. In this process, some absolute color information (the color of the light falling on the retina) is lost. Importantly, the loss is permanent, that is, no amount of introspective reflection can recover the direct experience of this information, which can only be reconstructed by artificial measures or viewing conditions.

The cascade continues. In the visual cortex, simple, complex, and hypercomplex cells detect bars at particular orientations, moving bars or lines, and the ends of edges ('edge stops') (Hubel and Wiesel 2005). Action potentials from each type of detector encode this information, a physical correlation between a stimulus in the environment and waves of ions crossing axon membranes. Those micro feats of detection constrain the detection of motion, shapes, color, etc. and eventually lead to recognition of houses, faces, tools, and Halle Berry (Quiroga, Reddy et al. 2005). But such high level recognitions do not efface their modest underpinnings. I see a house and at the same time I see the shape and color of the roof, comprising four edges at their specific orientations. I see the house in a configuration of parts and the parts in an arrangement suitable for a house. Top-down signals modulate these features. At one moment, I'm especially attentive to color, at another to texture, and so forth. These amplifications have their phenomenal manifestations as well.[23]

Hearing has its phenomenal cascade, too. Just now, my writing is interrupted as I perceive that the cat has pushed the bag of cat food off the counter in the kitchen. That's what I heard, the content of consciousness in the moment. But this lofty realization does not erase the basic sonic features that support the interpretation of the event. The sound of the impact of two pounds of crunchy nuggets was brief, inharmonic, ragged, and with a specific loudness and direction. The collected consciousness of the event is the collaboration of a cascade of sophisticated neural processes. Ultimately, I'll act on the highest-level conceptual content, but its humble components are copresent in awareness nonetheless.

In all these examples, cascading constraints appear as cascading aspects of consciousness. The cascade condenses in consciousness into single episodes of awareness, with all the richness of lived experience. The process remains subjective and objective at once. We sense the real world, but always from a point of view. We live among facts refracted through ambience and sentience.

3. THE ROAD AHEAD

To an optimist, it might seem like human neurophenomenology is almost complete. After all, several big phenomenological distinctions emerge as ecological-neural differences. However, there remains ample space for skepticism. The skeptic's hunch is some variation on: All this is not enough. At this point, philosophical reflexes engage, and the skeptical hunch gets translated into '*In principle*, all this is not enough,' i.e., no further elaboration will close the (remaining) explanatory gap. The optimistic materialist will be tempted to enter this debate on first principles. But she need not. Both optimists and skeptics can agree that the science is unfinished. It may be worth a side trip to show just how much remains to be done.

The 'standard picture' of the sensory cascade from transducers to embodied meaning emphasizes similarities between sight and hearing. This is the cascade from receptors to thalamus (vision targets the lateral geniculate nucleus, while hearing passes through the medial).[4] Both kinds of input then land in their primary sensory cortices. There and elsewhere the brain displays one of the great fascinations of cognitive neuroscience, namely, mapping. The primary visual cortex processes a highly modified and enriched visual (retinal) map. The auditory cortex, meanwhile, deploys a tonotopic mapping, an enriched projection of the basilar membrane. This seems to suggest that the two senses are strongly analogous. Meanwhile, the discussion above has drawn several contrasts between the senses. We have noted obvious differences in the behavior of ambient energies, in the structure and sequence of environmental features to be detected, in the function and arrangement of receptor sheets, in the information available for neural processing, in the computations required to make sense of the sensory stream, and in the experienced phenomenologies of the two sensory worlds. The deep phenomenological differences and the broad physiological similarities don't connect.

The prima facie conclusion is skeptical: the standard model fails to explain the conspicuous contrast between the senses. This explanatory shortfall is also apparent in the sheer numbers of neurons employed at the waystations of sensory processing. Figure 10.1 plots estimates of the number of axon fibers or neurons at early stages of the two sensory paths. The y-axis is logarithmic by necessity, as each step from the periphery involves orders of magnitude *increases* in involved neural resources.

For example, for both senses the neurons in the primary sensory cortices are approximately one hundred times more numerous than in the Geniculate. That leap in computing power undermines our confidence that topographical mapping explains very much of the processing of the primary cortices, since a one-to-one mapping from thalamus to cortex would require just 1% of the cortical resources. This glass is 99% empty. From the primary sensory cortices to the cortex, overall, the story is similar, with another hundred-fold jump in sheer numbers. At each stage, what are the other 99% *doing*? There are partial answers, involving the computation of higher-order properties of the topological/tonotopic maps, but the blunt takeaway is that a great deal still remains mysterious. In light of this rapid neuronal expansion, the dissimilarities between the senses have plenty of currently unknown resources for their support. That is, at each stage, there may be topographical mapping, but also so much more, and in that remainder the computational processes might be radically different.

Indeed, the computational distinctions between the senses, on top of their ecological differences, are arguably beyond the resolving power of existing

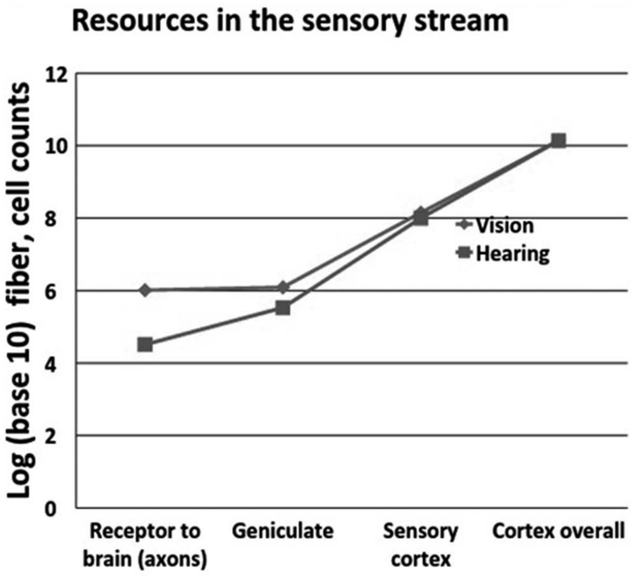

Figure 10.1 Neural resources in the sensory stream, at different stages in sensory processing. Showing log (base 10) counts of receptor to brain axons in the optic and cochlear nerves, neuron counts in Lateral geniculate nucleus (vision) and Medial geniculate nucleus (hearing), neuron counts in Primary visual cortex (Area 17) and Primary auditory cortex (Area 41), and neuron counts in the cortex overall. (Balazsi et al. 1984, Blinkov and Glezer 1968, Leuba and Kraftsik 1994, Spoendlin and Schrott 1989).

neuroscience technology. For example, fMRI might isolate intermediate stages that are combined in figure 10.1 (as 'cortex overall'). Imagine that we have a cortical map of areas of activation, vision vs. hearing. The discussion above implies that this degree of localization will not specify the distinctive processing that lends the senses their distinct phenomenologies. Similar points apply to EEG, MEG, and single-cell recording. Only at the circuit level might the distinctions be explained. Moreover, if the distinctions cascade, we will need the full input-to-output circuit to explain them. This is the ambition of several current projects, but Figure 10.1 also suggests how immensely daunting this project will be.[5]

Nonetheless, the glass is at least 1% full. Contrasting the two senses as embedded physical systems makes each pop out more clearly. As discussed above, the starting point is ecological, and then transducer-based. With both hearing and seeing, the physical distinctions are reflected in the phenomenology. Returning to blends and harmonies, we might be struck that we *can* explain why colors blend and sounds harmonize, a *phenomenological* distinction. Indeed, the explanation follows straightforwardly from the anatomy and physiology of the receptors themselves. Whatever else the 99% are doing, this proximal receptor-level distinction endures throughout the cascade of consciousness. So far, all of the fundamental phenomenal differences between seeing and hearing follow from neural distinctions.

The phenomenal distinctions between the senses are huge. If we knew nothing of the ecology of terrestrial light and sound, and nothing of early sensory processing, we might regard these differences as reflecting an ineffable, private, nonphysical, unique essence for each sense. This befuddlement evaporates as the straightforward facts of environmental energies and their transduction are disclosed. As these correlations between brain and experience pile up, slowly the burden of proof swings from the materialist to the anti-materialist. Each incremental increase in the understanding of human physical and physiological systems is at the same time a bit of phenomenology waiting to be translated from the objective language of action potentials to the subjective language of sights and sounds. This is *translation* from one description to another, where both denote one same underlying reality. There is no magic threshold where the spikes cease and consciousness begins. The roadblock here is complexity, not metaphysics. "Back to the things themselves," advised Husserl (Husserl 1900–01/2001, 168). Although Husserl had his doubts about the empirical sciences, we can nonetheless take the slogan very literally. Start with the world, and the mind will emerge in its natural role. The science so far may be miniscule compared to the mystery remaining. But the research in play is normal science (Kuhn 1962/2012), under the umbrella of versatile paradigms of materialism/biology/neuroscience. No stubborn anomaly perplexes the authors or readers of *Nature Neuroscience*. The message from phenomenology to science is simply this: Full speed ahead.

4. CODA: MUSIC IN THE MIDDLE

This analysis of two sensory modalities clearly presupposes that sensory systems are tools that animals deploy to actively probe their environments. The probing takes different forms for different senses, and, to specific senses, particular features of the environment are detectable. Evolution has operated for eons in an information-rich environment similar to that described in section I, and all the senses now seem to be optimized for that environment. Nonetheless, as animals probe the environment, they change it. Humans are particularly prone to this, creating "transformative technologies," artifacts with a pervasive impact on human life and society (Patel 2008, 400). These technologies leverage broad new affordances and ultimately change the way people think. Writing is an example. A new arrival from Mars would characterize our environment as saturated with words. Text underwrites most of the functions of civilized life, and literacy modifies the cognitive capacities of humans (Ong 1991). Nonetheless, writing arose too recently to shape the human genome, so its form and function must interleave with existing human physiology, including the various capacities of the sensory systems. The phenotype is the steady platform supporting the transformative technology.

Another technology that permeates our world is music. Every culture makes music, and the earliest musical instrument found so far is a flute made 43,000 years ago (Higham et al. 2012). Presumably singing has been practiced even longer. These features suggest that music is a transformative technology, like writing or fire. But unlike writing, fire, crop domestication, etc., it is possible to imagine civilization without music. Such a world might be boring, but food, clothing, and shelter don't strictly require a sing-along. So what is music *for*? Why is it ubiquitous in world cultures? There are many proposed replies. Here, we begin with the phenomenology, building on the contrastive worlds of seeing and hearing.

Music varies across cultures; one person might find another's music to be incomprehensible. But even on first hearing, a sequence of sounds will be heard as music (or not). 'Music' names a broad but recognizable entity; sound sequences lasting from seconds to hours can be heard as single musical productions. Because music stands out among sounds, producing or hearing music creates distinct experiences through distinct sensory processes. Thus music can be revealing of the capabilities and constraints of hearing and sensation overall.

As with hearing in general, the phenomenology of music begins with the ecology of musical sound, an anatomy of the artifacts of this potentially transformative technology. As discussed in section 1, hearing parses undifferentiated pressure waves into separate sound events. Some sounds are isolated, single bursts from a source (like the clink of spoon on cup), while other sounds can be organized into distinct streams (like speech, separated from the background hum of traffic). Ordinary audition begins with the 'primitive' sensory operations of grouping and stream segregation, in order to distinguish and locate sound sources in a complex auditory environment (Bregman 1990). Normally, hearing is dedicated to getting the source details right: what happened, where, and what's next. Music subverts these processes of grouping and segmentation. Separate musical sounds combine both 'vertically' (synchronically) and 'horizontally' (diachronically). Vertically, multiple distinct sound events form new composite unities, their combination heard as consonant or dissonant chords and harmonies (Sethares 2005). Over time, these composite events are assigned to one or more melodies, separate but related auditory streams. Ultimately, a number of streams are integrated in a complex whole, with a specific beginning and ending. Unlike non-musical streams, segregation is not based on identifying a single source. Rather, composite, distributed, asynchronous, and heterogeneous sources bind together in unified objects of perception.

Musicians and composers deliberately undermine the accurate segregation of sound sources. From sea chanties to symphonies, musical productions are auditory streams that are conspicuously *unnatural*. Music is an artifact through-and-through and flaunts its artificiality. Bregman and Woszczyk write:

> Music is *auditory fiction* in which the sounds of voices or instruments are combined to produce sounds that never appear in nature. The goal of music is often to temporarily lose the timbre and continuity of individual sound sources in favor of what can be called "ensemble blend qualities." (Bregman and Woszczyk 2004, 34)[6]

All of this is sonic deception. But where music is the percept, no one seems to mind. Makers aren't seeking to specify events at specific locations, nor are listeners concerned to reconstruct those events. In this respect, musical audition is detached from ordinary, 'natural' hearing. (See also Clarke 2011, 22)

Moreover, the individual components of musical productions are not found in the natural world. Different cultures have developed different 'musical systems,' heuristic constraints that shape the making and hearing of music.[7] At the root of nearly all of them are *stable tones*, persistent sounds at steady pitches within pieces (Powell 2010). These tones are almost always drawn from a limited tone lexicon—a *scale*, which will be different for different compositions/performances. In Western music, these are the major, minor, and modal scales. Some Asian traditions use pentatonic scales, while others use scales with different intervals but typically five to seven distinct tones. The exact pitches vary, but the intervals between pitches are stable. All of these world musics replicate pitch classes across octaves, doubling (or halving) the frequencies of the selected scale. Tones in use in a particular performance/composition are also constrained by heuristics of *harmony*. The heuristics include both horizontal constraints (governing melodies and their variations) and vertical constraints (governing simultaneous tones, i.e., chords). Finally, tones and their combinations appear in temporally periodic sequences, or *rhythm*. The time intervals between music events in a given rhythm are relatively stable (Patel 2008).[8]

The building blocks of musical systems constrain musical properties from moment to moment, but productions are further limited by another feature unique to music, namely, self-similarity over time. Self-similarity is built into the heuristics. At every scale, music creates recognizable sound patterns that extend in time. As mentioned above, individual musical sounds endure longer than most natural sounds. Other properties of productions, like timbre, key, and meter, endure longer still. But self-similarity reaches further than the sway of the basic heuristic constraints: musical phrases repeat throughout almost all productions. Ollen and Huron (Huron 2006, 228) examined a cross section of world music melodies and discovered that on average 94% of musical units of two seconds or longer will repeat at least once within any piece of music (see also Lloyd 2011). The collective effect of the constraints of musical systems is to create a soundscape that is predictable over time. These properties conspire to link musical sounds into coherent sound streams, continuous and shifting but bound in a single phenomenal entity. The interest in music lies in the interplay of novelty and expectation (Huron 2006), but musical surprise is only possible against a background of expectations, which are shaped by the stability of many of the musical properties we've surveyed.

As a stream of sounds, a musical stimulus is repetitive. But as an experience, repetition involves recognition of repeated material at every time scale. Musical experience involves multiple encounters with the same sound (or sequence), re-identified. This continuous return to the same pattern of sounds, at every time scale, shapes musical experience into forms distinct from non-musical, event-centered hearing. However, this patterning of sensation is not utterly novel. Instead, it is the pattern typical of *seeing*. Revisit that now-empty cup of coffee and recall the distinctive features of visual phenomenology that so sharply separated the world of sound from that of light. The continuous light energy bathing a typical visual landscape created a stable scene. Every lit corner of the landscape streams information in all directions, and the eye can sample the stream either briefly or at length. Each sampling yields a relatively steady mix of light frequencies, analogous to the stable tones comprising any musical production. Visual exploration affords resampling of any part of the illuminated scene, or repeating any sequence of glances. Music, with its relentless self-similarity, affords a similar multi-scale repetition.

The world of visible objects exhibits stability overall, affording visual themes and variations—the general phenomenological structure of music as well. In addition, harmony describes at least two features of the visual world and its exploration. Spatial relations are stable in the visual field, analogous with synchronic/vertical harmony. Music also models the interaction of organism and environment. Husserl describes the interaction of action and perception, describing the 'laws' by which the proprioceptive awareness of movement counters the shift of sensory information from a stable source. For example, as my eyes pivot right, the retinal image slips to the right as well. But we don't perceive a scene sliding to the left, because the afferent awareness of the shifting gaze is the exact counterpoint of the shifting scene. Husserl names this counterpoint of afferent and efferent a "harmony of desire." (Husserl 1973/1997, sections 29, 42, and similarly sections 30, 33, and 54; see also O'Regan et al. 2004). Music makes the metaphor literal.

In short, vision constitutes objects that are spatially extended, in a spatial environment. Music mimics this, creating sound objects that extend in time. These sound objects borrow ecological properties from the information landscape of vision. Music embodies the actions of visual perception, but in a musical landscape with musical objects. Music thus enacts visual thinking. It creates stable sound objects, and in its repetitions and variations mimics the patterns of exploration of a stable, illuminated visual world.

So far this essay has avoided speculating about the ‘meaning of music.’ But real phenomenology bears on this question. Most conceptual and theoretical discussions of music use language as a foil, and particularly speech. Both speech and music are sonic artifacts, and both exhibit a combinatorial, generative syntax (Lerdahl and Jackendoff 1983), but the resemblance ends there. In brief, speech relies mostly on timbre, not pitch, for distinguishing meaningful sounds. Even in tone languages, speech tones are not drawn from scales with fixed intervals. (Tone languages use relative intervals; the exact pitches and intervals of meaningful sounds can change within and between utterances Patel 2008, 44–45). Temporal properties matter for spoken language but these are not regular; and harmony has no role at all (ibid., 141ff).[9] Perhaps the deepest difference is semantic: music simply cannot denote concrete referents. In that regard you cannot *say* anything with music (apart from lyrics). Some philosophers respond to this aporia by seeking new referents that accommodate the vagueness of musical reference. As a result, the philosophy of music reconstrues musical representation as some combination of the abstract, the purely emotive, and/or inarticulate bodily movement (Kivy 2002). In contrast, the present analysis locates music in a phenomenological space *between* hearing and seeing. Thus, music augments a landscape of contingent, punctate events. It creates something different in the landscape of sound.

For thousands of years our human ancestors faced an urgent world that left little time for philosophizing. The immediacy of experience in that pre-technological era offered little need or opportunity (or words) for abstract reflection. Nonetheless, we can imagine the first songs resonating around the daily business of survival. If these songs had the features still audible in world music today, their singers would have augmented their sensory world with a simulacrum of enacted perception itself, creating a model of action in a stable world out of the most unstable of materials, sound. Their music, and ours, creates a phenomenal space in which sound mimics the objective world of sight. Aldous Huxley has written that “[a]fter silence, that which comes nearest to expressing the inexpressible is music” (Huxley 1931, 17). A great deal of human experience evades verbal description—expressing that nebulous dimension of the lived world has been repeatedly assigned to music. These various philosophies of musical expression are not necessarily exclusive. Emotional expression and abstract representations of the dynamics of sentient life are arguably all part of the function of music. This discussion has added one more possibility: music rings the changes of active exploration of a stable environment. It embodies in its own flow the relationship of sensory events in a flow of many simultaneous layers. Music may have arisen and been shaped by an implicit, enactive, awareness of the dynamics of sensory experience itself, as it occurs in a mobile, inquisitive animal in a relatively stable environment. In that way music can be a reflection—perhaps the first and oldest echo—of the human situation in the world. Before and beyond speech, it may be pre-reflective reflection, the wordless sound of philosophy itself.

NOTES

1. Gibson again: "An important fact about the affordances of the environment is that they are in a sense objective, real, and physical, unlike values and meanings, which are often supposed to be objective, phenomenal and mental. But, actually, an affordance is neither an objective property nor a subjective property, or it is both if you like. . . . It is equally a fact of the environment and a fact of behavior. . . . An affordance points both ways, to the environment and to the observer" (1979, 129).
2. So, action potentials appear *as* edges, motion, and so forth. One trap of these debates is to reify mental states as distinct entities and then ask how a neural spike could cause a thought, making one event (with two descriptions) into two events. This is like asking how the evening star causes the morning star. This is a hard question, but only because it is nonsensical.
3. In normal environments the cycle of bottom-up and top-down is unimpeded, and so information propagates phenomenally as in the examples here. Special experimental conditions (without ecological parallel) can block the flow. When that happens information is lost (i.e., fails to propagate) and thus ceases as part of experience as well.
4. This elides waystations prior to the geniculate. Signals along the auditory nerve diverge at the cochlear nucleus and project in parallel pathways to the superior olive, lateral lemniscus, and inferior colliculus—and then to the thalamus. The medial geniculate arguably handles a signal that is already enriched in ways the visual signal is not. For an overview, see Kandel et al. (2013).
5. The projects include: The Human Brain Project, https://www.humanbrainproject.eu/; The Connectome Project, http://www.humanconnectomeproject.org/; 'Blue Brain,' http://bluebrain.epfl.ch/; and The BRAIN Initiative, http://www.nih.gov/science/brain/
6. The converse also occurs when a melodic line jumps rapidly between high and low notes, creating the illusion of two or more melodic lines from a single source—otherwise known as implied polyphony or melodic segregation, frequent in Baroque music (Bregman 1990, 464).
7. As an art form, Western art music has undergone the sprawl of modernism, but the avant garde creations of Schoenberg, Stockhausen, Cage, et al. are understood in reference to schemas of musical production familiar to audiences. These works are specific rejections of aspects of musical traditions.
8. In most of these features, human musical productions are unlike quasi-musical signaling by other animals. In general, the songs of nonhuman animals are stereotyped in many respects: who sings, when, and for what (adaptive purpose). Bird and whale song display limited variation in sequence, apparently lacking the versatile combinatorial syntax of human music (Patel 2008, 355, 356).
9. A striking demonstration of the difference between music and speech is Diana Deutsch's 'speech to song' illusion. As a segment of a sentence is repeated in a recorded loop, slowly the perception of words faces, replaced by a vivid melody of speech tones, a song. When the looped phrase is reinserted in its original context, one hears a sentence with a burst of song embedded in it (Deutsch et al. 2011). Steve Reich's *Different Trains* is a beautiful example.

REFERENCES

Balazsi, A.G., J. Rootman, S.M. Drance, M. Schulzer and G.R. Douglas (1984) "The Effect of Age on Nerve Fiber Population of the Human Optic Nerve," *American Journal of Ophthalmology* 97 (6): 760–66.

Blinkov, S.M. and I.I. Glezer (1968) *The Human Brain in Figures and Tables: A Quantitative Handbook*, New York: Basic Books.

Bregman, A.S. (1990) *Auditory Scene Analysis: The Perceptual Organization of Sound*, Cambridge, MA: MIT Press.

Bregman, A.S. and W. Woszczyk (2004) "Controlling the Perceptual Organization of Sound: Guidelines Derived from Principles of Auditory Scene Analysis" in K. Greenebaum and R. Barzel (eds.), *Audio Anecdotes: Tools, Tips, and Techniques for Digital Audio*, Vol. 1, Natick, MA: A.K. Peters, 33–61.

Churchland, P.M. (2005) "Chimerical Colors: Some Phenomenological Predictions from Cognitive Neuroscience," *Philosophical Psychology* 18 (5): 527–60.

Clark, A. (2008) *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*, Oxford: Oxford University Press.

Clarke, D. (2011) "Music, Phenomenology, Time Consciousness: Meditations after Husserl" in D.C. Clarke and E. Clark (eds.), *Music and Consciousness: Philosophical, Psychological, and Cultural Perspectives*, Oxford: Oxford University Press, 1–29.

Deutsch D., T. Henthorn and R. Lapidis (2011) "Illusory Transformation from Speech to Song," *Journal of the Acoustical Society of America* 129 (4): 2245–52.

Gibson, J.J. (1979) *The Ecological Approach to Visual Perception*, Boston: Houghton Mifflin.

Heidegger, M. (1927/2008) *Sein und Zeit*, Tübingen: Niemeyer; *Being and Time*, New York: Harper Perennial / Modern Thought.

Higham, T., L. Basell, R. Jacobi, R. Wood, C.B. Ramsey, and N.J. Conard (2012) "Testing Models for the Beginnings of the Aurignacian and the Advent of Figurative Art and Music: The Radiocarbon Chronology of Geissenklosterle," *Journal of Human Evolution* 62 (6): 664–76.

Hubel, D.H. and T.N. Wiesel (2005) *Brain and Visual Perception: The Story of a 25-year Collaboration*, New York, NY: Oxford University Press.

Hull, J.M. (1992) *Touching the Rock: An Experience of Blindness*, New York, NY: Vintage Books.

Huron, D.B. (2006) *Sweet Anticipation: Music and the Psychology of Expectation*, Cambridge, MA: MIT Press.

Hurvich, L.M. (1981) *Color Vision*, Sunderland: Sinauer Associates.

Husserl, E. (1900–01/2001) *Logische Untersuchungen*, Halle: Niemeyer; *Logical Investigations*, trans. J.N. Findlay, Dermot Moran (ed.), London: Routledge. Translation of second edition, 1913.

———. (1928/1964) *Vorlesungen zur Phänomenologie des inneren Zeitbewusstseins*, hrsg. M. Heidegger (ed.), Halle: Niemeyer; *The Phenomenology of Internal Time-Consciousness*, trans. J. Churchill, Bloomington: Indiana University Press.

———. (1973/1997) *Ding und Raum, Vorlesungen 1907*, hrsg. U. Claesges (ed.), Husserliana Band XVI, The Hague: Nijhoff; *Thing and Space: Lectures of 1907*, trans. R. Rojcewicz, Boston: Kluwer.

Huxley, A. (1931) *Music at Night, and Other Essays*, Garden City, NY: Doubleday, Doran & Company, Inc.

Idhe, D. (2007) *Listening and Voice: Phenomenologies of Sound*, Albany: State University of New York Press.

———. (2012) *Experimental Phenomenology: Multistabilities*, Albany: State University of New York Press.

Kandel, E.R., J.H. Schwartz, T.M. Jessell, S.A. Siegelbaum and A.J. Hudspeth (2013) *Principles of Neural Science*, New York: McGraw-Hill.

Kivy, P. (2002) *Introduction to a Philosophy of Music*, Oxford: Clarendon Press.

Kuhn, T.S. (1962/2012) *The Structure of Scientific Revolutions*, Chicago: The University of Chicago Press.

Lerdahl, F. and R. Jackendoff (1983) *A Generative Theory of Tonal Music*, Cambridge, MA: MIT Press.

Leuba, G. and R. Kraftsik (1994) "Changes in Volume, Surface Estimate, Three Dimensional Shape and Total Number of Neurons of the Human Primary Visual Cortex from Midgestation until Old Age," *Anat Embryol (Berl)* 190 (4): 351–66.

Lloyd, D. (2011) "Mind as Music," *Frontiers in Psychology* 2 (63).

Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, San Francisco: W.H. Freeman.

Mensch, J. (2014) "A Brief Account of Husserl's Conception of Our Consciousness of Time" in V.L. Arstila and Dan Lloyd (eds.), *Subjective Time*, Cambridge, MA: MIT Press, 43–60.

Noë, Alva (2004) *Action in Perception*, Cambridge, MA: MIT Press.

O'Regan, J.K. (2001) "A Sensorimotor Account of Vision and Visual Consciousness," *Behavioral and Brain Sciences* 24 (5): 883–917.

O'Regan, J.K., E. Myin and A. Noë (2004) "Towards an Analytic Phenomenology: The Concepts of 'Bodiliness' and 'Grabbiness' " in A. Carsetti (ed.), *Seeing, Thinking and Knowing*, Dordrecht: Kluwer Academic Publishing, 103–14.

Ong, W.J. (1991) *Orality and Literacy: The Technologizing of the World*, London: Routledge.

Patel, A.D. (2008) *Music, Language, and the Brain*, New York: Oxford University Press.

Powell, J. (2010) *How Music Works: The Science and Psychology of Beautiful Sounds, from Beethoven to the Beatles and Beyond*, New York: Little, Brown, and Co.

Quiroga, R.Q., L. Reddy, G. Kreiman, C. Koch, and I. Fried (2005) "Invariant Visual Representation by Single Neurons in the Human Brain," *Nature* 435 (7045): 1102–7.

Sethares, W.A. (2005) *Tuning, Timbre, Spectrum, Scale*, London: Springer.

Spoendlin, H. and A. Schrott (1989) "Analysis of the Human Auditory Nerve," *Hear Res* 43 (1): 25–38.

Varela J.F., E. Thompson and E. Rosch (1991) *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge, MA: MIT Press.