

## 12 A Preliminary Phenomenology of the Audiobook

*D. E. Wittkower*

Philosophy has been called “a radical asking of the common questions of everyday.”<sup>1</sup> Here, our object of inquiry is not one as fundamental to experience as those typical of philosophical inquiry, and our question not quite so common or everyday. Rather than radicalizing general questions such as “What is that?” “Who are you?” or “What time is it?” we will instead ask two interconnected questions:

- What is it like to listen to an audiobook?
- What are we listening to when we listen to an audiobook?

Before turning to the experience of listening to an audiobook, it is worth making a few preparatory notes on the material conditions necessary for the possibility of this experience. Wired or wireless, the audiobook can reach us only through our involvement in a larger technical system, both in the preparation of the experience in question and in the circumstances of the experience itself. The preparation necessary for the possibility of the experience, completed within a larger technical system, may differ in each instance, but a typical preparation might include obtaining an audiobook on CD or through download, adding the audio files to a digital library on a desktop computer, adding the files within a playlist synced to an MP3 player, and updating the MP3 audio library. The circumstances necessary for the possibility of the experience itself will differ as well but invariably include integration in one of a number of larger technical systems, such as listening to the audiobook on a car stereo while driving, on a home stereo while performing domestic tasks, or on an MP3 player while walking or at the gym.

We must also note that the activity is neither isolated nor all-consuming: in fact, if we are to understand the experience of listening to an audiobook, we should not assume that the convergence of intermixed and simultaneous activities has no bearing on the phenomenology of listening, or that coincident activities would not have different effects on the experience. Thus, in a full phenomenology of the audiobook, we would have to consider listening

as an experience sufficiently situated and specific to be attuned to the influence of convergent activities but general enough to be able to provide a resonant shared notion of a ‘what it’s like’ for audiobook listening.<sup>2</sup> This most general understanding of the experience of audiobook listening exhibits many distinct although interrelated movements:

1. The audiobook is a temporal object of experience;
2. The audiobook is spoken;
3. The audiobook has a speaker;
4. The audiobook is started, stopped, and restarted;
5. The audiobook forms a context of physical and social experience.

Owing to space limitations, I will concentrate in this chapter on the audiobook as a temporal object of experience, as this is directly relevant to phenomenological discussions of aural experiences originating with Edmund Husserl. The remaining movements will then be addressed more briefly, and the reader welcomed to consider these commentaries as spurs, notes, or proposals for future research.

## 1. THE AUDIOBOOK IS A TEMPORAL OBJECT OF EXPERIENCE

The audiobook is audible and auditory. This will obviously distinguish it from the written work of which it is a performance, and the specific time-boundedness of the auditory object must be addressed in order to make sense of what we are listening to when we listen to an audiobook. At the same time, while we will compare the audiobook to previously theorized auditory objects of experience, such as music or the speech of a present other, we must distinguish it from these other objects as well.

“All sensation takes place in time,” Walter Ong has noted, “but sound has a special relationship to time. . . . There is no way to stop sound and have sound. I can stop a moving picture camera and hold one frame fixed on the screen. If I stop the movement of sound, I have nothing.”<sup>3</sup> Even the film, though, does not really exist as an object of experience within its still frames but is essentially in the movement and flow from one moment to the next. As Maurice Merleau-Ponty put it, “[A] film as a perceptual object . . . is not a sum total of images but a temporal gestalt,”<sup>4</sup> and the perception of its “movement,” as with the perception of a melody rather than individual notes, is really the primary and fundamental content of our experience of these objects:

melody does not perceptibly change when transposed, that is, when all its notes are changed while their interrelationships and the structure of the whole remain the same. On the other hand, just one single change

in these interrelationships will be enough to modify the entire make-up of the melody. Such a perception of the whole is more natural and more primary than the perception of isolated elements.<sup>5</sup>

So too, as Don Ihde has put it, “[W]hen I listen to someone speak, I do not ordinarily hear a syllable at a time, or even a word, but I hear the larger melody and flow of speech as an ongoing rhythmic unity.”<sup>6</sup> The auditory speech-object is the process of movement between sounds and phonemes. Hence a fuller discussion of how we hear a melody rather than individual notes will allow us to give a “close listening” to the experience of hearing a sentence rather than tones, consonants, and phonemes.

### How We Hear a Melody

Each particular sound—say, the note of middle C—is physically present only while it is being actively created by the vibration of the instrument or speaker system. As we experience the C followed by an F, and then a G, a C major tonality is established in the melody. Should the G be followed by a B, we would experience the B as leaning upward toward a return to the tonic, while, had we simply heard a B in isolation, we would experience no such tension. Furthermore, in order to have an experience of the B as a leading tone, we must not merely remember the preceding tones but must continue to feel them as well—but we must feel them as present under erasure for, as John Brough points out, “a melody, whose successive notes were heard all at once as now, would not appear as a melody at all, but as a crash of simultaneous sound.”<sup>7</sup> No longer foreground but not yet absent to perception, the notes that are no longer physically present to the eardrum must be phenomenally present to the listener in order to form the background against which the primary impression of the sounding note appears. It is only this presence-undererasure of the absent which makes it possible to hear a melody rather than a series of mere tones or a clashing cacaphonic muddle.

Husserl describes this as retention—sharply distinguished from memory. Retention is a kind of primary remembrance in which we maintain what just was as a continuing presence to experience, which relativizes the ongoing primary perception of the senses. Memory as we usually use the term is not the presence of a still-perceived ‘just-having-been,’ but rather a prior ‘now’ which appears to us only through our recall of it—in Husserl’s words, “*not perceived, i.e. self-given, but presentified*.”<sup>8</sup> Perception of objects—such as the melody—which appear in several parts, only one of which is primally present at any moment, requires retention, not memory. Merleau-Ponty makes the same point in a different way:

If the past were available to us only in the form of express recollections, we should be continually tempted to recall it in order to verify its existence . . . whereas in fact we feel it behind us as an incontestable acquisition.<sup>9</sup>

The presence of a perceived object of experience includes, in addition to a 'now' and a 'just-having-been,' a primordial unity with a future as a 'just-about-to-be.' In the prior example, the experience of the B as a leading tone requires not only a felt carrying-along of the 'just-having-been' but also a projective hearing in which the C that is not heard is experienced in its absence within our hearing of the B. The presence to experience of the B as a tone pushing us toward a return to the home tone of C is *constructed* by the presence of the having-just-been but *consists of* the presence to experience of the just-about-to-be: the C which is not heard, or which is 'heard' as not-yet-present. In harmonies this is felt even more clearly: the movement I–IV–V–VI<sup>7</sup> is experienced as a false resolution precisely because the VI<sup>7</sup> chord is heard as superimposed against the tonic chord established by the dominant chord as an expectation. The VI<sup>7</sup> chord sounds false to the ear not upon reflection or as an intellectual judgment, but rather it is felt as false from the moment the tones sound because it is heard as failing to match up with the tonic chord present to experience at that moment as the anticipated resolution that we are precisely *not* listening to. This presence in the 'now' of an expected future is discussed by Husserl as *protention* and is the futural equivalent of retention, as Thomas Clifton has explained:

Protention is the term for a future which we anticipate, and not merely await. Awaiting, like recollection, implies a disengagement from the present, whereas, experientially, the now which we perceive is colored by the way we intend the future. Intending a future with respect to a given event means to attach significance to that event in proportion to the way the present and future are attached to, yet distinguishable from, each other.<sup>10</sup>

In listening to speech, too, we certainly do not merely await to hear what words will "arrive" next but instead actively intend the future in modified ways from one phoneme to the next. This is clearly the case when listening to the speech of a present other, as the future is projected forth by context, relationship, body language, and so forth, but this structure of active anticipation and construction of meaning is also present in a full and vibrant way while listening to recorded speech. As we will see, when we listen to the audiobook, we hear every word as along with the words that have been and as against the words that could have been, just as surely as, in the melody, we hear each note with the notes which preceded it and against the notes which might have followed.

### How We Hear a Sentence

Processes of retention and protention are necessary for the listener's comprehension of the spoken word: each new phoneme is heard as relativized by the phonemes preceding it and construct for us a protended anticipation of the words likely to follow. The adjective alerts the listener to a

substantive to follow; the pronoun projects forward a verb. The listener actively constructs the meaning of sentences as the phonemes are sounded and replace one another through time: in order for the subject to be heard as the subject, the listener must hear it along with the not-yet-presence of the predicate which it, as subject, implies will follow. The predicate only appears to the listener as a predicate insofar as the listener retains the subject of the sentence as the just-having-been, which is the background upon which the predicate appears; the predication of the predicate inheres in the absent presence of the retained subject.

All of this is true of the written word as well but in differing ways. The written word is put into motion, so to speak, by the action of the reader. The reader moves forward across the page, replacing the phonemes present within the 'now' of the inner voice at her own pace, circling back to reread as necessary. The spoken word, when containing clauses and conditionals, must be heard by the listener as a series of related points, each of which must be retained by her and actively reconstructed as the succession of phonemes continue to bring further meaning to earlier components. Consider this sentence as it would have to be understood if spoken:

Consider (imperative action, protends an object) this (protention prolonged: what is it?) sentence (protended object supplied; imperative must be retained in order to contextualize the intended encounter of this object) as (modifies retained imperative action "consider") it (pronoun calling for retention of substantive "this sentence") would (clarifies retained modification "as" of retained imperative "consider"—the listener now knows we are going to consider the sentence "as" something in some counterfactual form yet to be determined) have (the listener hears /æ/ rather than /ə/, signaling that this "have" does not primarily modify the "would"—as in the spoken "would've," which uses a schwa—but will instead modify something which is to follow. If the speaker is American, this word will be pronounced /hæf/, indicating that this "have" is not the "have" of possession, which is /hæv/ in both UK and US English, but will instead form the idiomatic "have to" of necessitation) to (with retained "have" forms idiom of necessitation, protending a subject to be necessitated yet to be determined) be (infinitive verb fragment, the listener must still wait for more context) understood (that which is necessitated by the "have to" is supplied, and the retained phrases "have to" and "to be" can now be used to connect supplied content "understood" with object "sentence" under imperative "consider") if (indicates that the specific nature of counterfactualism, implied previously by "would," will now be supplied) spoken (retained sentence now placed under counterfactual conditional supplied).

Protention and retention are both different from conscious or intentional recall and expectation, and this breakdown of the process of listening is not

meant to represent the self-aware movements of the mind but instead meant to clarify and make present to awareness the unnoticed and unreflective process of structured listening and active contextualization which makes understanding possible. This structured listening constructs a number of simultaneous timelines. The subject is given, protending a predicate. But then, a modifying clause is given to the subject, and we must maintain our protention of predication while modifying the retained subject. Having circled back, we may again move forward in the construction of meaning. As each word replaces the last in the succession of 'nows,' we construct the meaning of the sentence through a nonlinear cycle of emerging protentions which are variously resolved into the retained content, and the overall meaning of the sentence emerges out of the circling back of elements upon one another, as we might see water as flowing downstream even though it does so through a complex and meandering process of moving up, down, and backward through ripples and eddies on the surface.

## 2. THE AUDIOBOOK IS SPOKEN

In the written word, the particular modes of relevance of one word upon another are not communicated through grammatical roles alone but also take place through the occult actions of punctuation marks. In the spoken word, similar signals are given through precisely timed pauses and changes in tone. The commas used to offset modifying or explanatory clauses, such as those surrounding this aside, are intended to be heard differently, and to construct meaning differently, than those within a list. The colon indicates content: that which follows it loops back upon and superimposes on that which precedes it. The modifying or explanatory clause—set off by commas or em dashes, like this—is lowered in pitch relative to the primary “timeline” of the sentence. Some words are emphasized in volume or enunciation in order to signal to the listener that these words are essential to the retrospective reconstruction of the meaning of the sentence in question. Words in subordinate clauses and noun phrases are run together subtly in order to indicate their unification as objects distinct from other “moving parts” within the overall claim. Parentheticals are uttered sotto voce (at half voice, like this).

Punctuation marks may seem far clearer than the pauses and vocal modulations of the spoken word, but it is not obvious that they have specifiable meaning. It seems just as likely that they notify us simply that something is supposed to happen at a given point in the sentence, and—like “reading” vowels in Hebrew—the reader is depended on to be able to tell based on context what is supposed to be happening. Today’s punctuation usages are formalized and do not always track the rhythms which lend speech its intelligibility. Turning back to earlier forms of written English, we can see punctuation which is less a formal feature of writing and more

an encoding of speech. For example, consider a passage from John Stuart Mill's *Utilitarianism*:

Against this doctrine [the Greatest Happiness Principle], however, arises another class of objectors, who say that happiness, in any form, cannot be the rational purpose of human life and action; because, in the first place, it is unattainable: and they contemptuously ask, what right hast thou to be happy? a question which Mr. Carlyle clenches by the addition, What right, a short time ago, hadst thou even to be?<sup>11</sup>

Whereas this passage closely tracks the affordances of speech in aiding the listener's active construction of meaning, it is still clearly *composed* and far different from speech as extemporaneously spoken.

~To speak with someone, that is, to speak with a particular, well to speak in the way that we really do talk most of the time, requires that there's a lot of back and forth and often (well, I do this—you probably do too), if you really pay attention to what someone says, we'll start a sentence three or four times before finishing it. And then start back up in the middle again! And when the speaker moves on from one point to another without finishing the first thought, maybe we know where she was going, or, if we don't, we forget about it because we're concerned with communicating with the person in front of us! The words are just a means to that end, and if there are a few loose ends at the end, who cares! (—and plus, there's no evidence, right?) Spoken words move their cargo like ants pulling along a bit of bread: the ants fall down, some drop it, others keep going, some peel off to go do something else, and some of them push it the wrong way, and little by little, the bread moves on down the line.~

The written word is fully available to the reader,<sup>12</sup> to be engaged with at her own pace and in the order and level of care that she prefers. The spatial presentation of writing allows for a nonlinear visual encounter because space, unlike the successive presentation of time, contains and co-presents its elements. The reader may circle back, repeat, review. And so, in contrast to the way spoken word functions between present conversationalists, a written thought must be completed and consistent before it is expressed.

In between the starting and stopping of playback, the audiobook follows its own rhythm and pace in a context-insensitive and user-independent way. Listening to an audiobook is not like listening to a person speak—persons speaking to persons make eye contact, pause to invite signs that the listener is paying attention, use body language to emphasize or change meanings, and um and hem and haw. The audiobook proceeds at the same inexorable rate while the audience drifts in and out of attention.

(Please read this paragraph backward from the end.) .paragraph this in simulate to tried have I which experience this is It .something missed we've find we if sentence the of beginning the to back look simply cannot we because ,end the at together all it piece to passed just has what of enough



remember to working actively ,procession their on carry words the while on going is what out work to attempt must We .sound of nature momentary the of difficulty inherent the exacerbates understanding and meaning create to take-and-give interpersonal the use to inability This .needed when back circle and listener the to respond to ability the without ,relentlessly forward continues narrator The .language written or spoken either of affordances the of many lacks audiobook The

The audiobook, as an audio performance of the written word, contains the determinate and preconsidered meaningfulness of the written word but without the affordance of nonlinear options for reader engagement. However, the richness of human speech seems to be sufficient compensation to support comprehension, even without the give-and-take of listening to a present other. Not only are the guiding rhythms of speech reintroduced, substituting for punctuation marks—their pale written equivalent—but additionally every word is made replete with meaning through its intonation. As Ihde has observed,

For the reader who comes upon the word on a page, the field and its unsaid significance is a dark obscurity. . . . But if this word is *spoken*, there is already a certain potential field and presence of unsaid significance in the voice. If “Adam” is said in an angry voice, imploringly, or in a quiet whisper, each sounded presence allows the “bare word” to emerge from some of its obscurity in the sounding of its presence.<sup>13</sup>

Similar to Heidegger’s observation that Dasein always finds itself having a mood (*Stimmung*), we may observe that the voice always has some kind of attunement. Just as “the pallid, evenly balanced lack of mood . . . is far from nothing at all,” so too is the word spoken with no inflection and a flat affect just as full of meaning in its performance as if spoken animatedly.<sup>14</sup> The voice cannot but carry with it an emotional component, even when that component is emotionlessness. Marshall McLuhan noted that “[t]here are not many ways of writing ‘tonight,’ but Stanislavsky used to ask his young actors to pronounce and stress it fifty different ways. . . . The written word spells out in sequence what is quick and implicit in the spoken word.”<sup>15</sup> The audiobook, although it contains this narrative description of what is emotively present in the spoken word, contains the replete modulations of the spoken word as well.

The pause, too, plays a great role in aiding the construction of meaning. The pause can function as a grammatical marker, signaling when to bind together meanings with preceding substantives, or when to expect further conditionals, modifiers, actions, and so forth. Pauses serve not only as signals to engage in these forms of synthesis and expectation but also as opportunities to do so. When the narrator moves along too quickly, or without suitably lengthy pauses at constitutive interstices, we very easily become confused and fall behind. The longer pauses which come at the end



of sentences, paragraphs, or sections serve a similar role, except that they signal the completion of a larger process of construction and provide the opportunity for the listener to bind together a larger set of meanings. As Husserl points out, these final pauses are not experienced as silence:

One speaks of the dying or fading away etc., of the content of sensation when veritable perception passes over into retention. Now, according to the statements made hitherto, it is already clear that the retentional “content” is, in the primordial sense, no content at all. . . . The retentional sound is not actually present but “primarily remembered” precisely in the now.<sup>16</sup>

Consider the natural pause which occurs at the end of a classical performance. It is surely possible for an audience member to applaud too soon, but this rarely occurs. Instead, the audience members seem on the whole to have an agreement about when the pause necessary to the piece has elapsed, or they simply await the downward return of the conductor’s baton. The duration of this pause is rightly regarded as a part of the composition, for applause too early is heard as an interruption despite the fact that sound is no longer being produced. Without this pause, the music stops but does not end; the binding together of sound into unitary temporal experience requires this pause, and thus there can only be an ending in the sense of coming to fruition when there is an appropriately timed pause which, as the ending of the piece, is as integral to the experience as any note sounded within it.<sup>17</sup> The appropriate duration is dictated by the mental reverberation required by binding the experience, not by a physical echo which may be present. It is likely that a haunting piece which trails off into a lonely and hopeless pianissimo would require a longer pause before applause than a louder piece with less emotional depth, even if the latter is performed in a very live room and the former in a more muted space. Unlike the physical reverberations of the echo, the mental reverberation of retentional meaning-binding in the pause grows louder as the pause extends. Or, at least, up to a certain point—certainly the pause can be too long, at which point the listener has finished attending to what has just been heard and has become distracted by other events, or the silence ceases to retain what came before and begins to pretend what is about to occur. In the concert hall, we await the drop of the conductor’s baton, and at some point we are no longer still ‘listening’ to the now-absent music but begin instead to think, “What’s going on? Should I clap? Maybe it isn’t over yet . . .”

In order to properly bring sentences, paragraphs, and chapters to fruition, the narrator of an audiobook must similarly pause for the right amount of time and do so for the same reasons. Should the pause be too short, comprehension is inhibited and the listener becomes confused or disengaged. Like the conductor holding her hand aloft, the narrator can signal to the audience not to move on, not to be distracted, and not to stop paying

attention. For this reason, it may be that audiobooks have special value to offer to poetry, for readers may have difficulty lingering properly over phrases, and the audiobook forces an appropriate reflective duration (even if the reflection itself does not always take place). As Nick Piombino put it, “[T]he aural ellipses of the contemporary poem ensure that there will be spaces for invention on the part of the listener. . . . this is all the more important in contemporary life where there is so much talking and so little listening.”<sup>18</sup> When we read the written poem, we are ‘talking’ and ‘listening’ to our ‘inner voice,’ and unless we make a serious effort to internally perform the poem, it is very easy to fail to hear the poem at all. Having the poetic work performed provides a structure to support the listener’s active listening: she must wait, and she may then be more likely to wonder and protend and reflect. This effect extends to poetic and dramatic passages in prose works as well.

Surely, here too the pause can be too long. We are trained by radio to be bewildered by “dead air” in a way that we are not in the concert hall, and when there is any somewhat long pause, we begin to be distracted by our surroundings or to ask whether the track has ended, whether something has gone wrong with the playback, or whether it’s time to insert the next CD.

Here, the amount which can be carefully and precisely covered within a single book chapter is at an end. In order to make some attempt at appropriately broad coverage of the chapter’s topic, as previously mentioned, I will make some briefer and more informal commentaries on other aspects of a phenomenology of audiobooks.

Thus far, we have seen how listening to an audiobook requires a listening which binds together past, present, and future in an active construction of meaning. This process, while immensely complicated, is one to which we are well accustomed from its similarity to communication with a present other. We have also seen how pauses and intonations strongly aid comprehension of complex written prose, not necessarily intended to be read aloud, even when the affordances of the written page are removed.

Taken together, we might expect these aspects of audiobook listening to result in a displacing or entrancing experience. There is much work to do and attention demanded, and there is a flow of future with past into present which follows no rhythm but its own—and all of this is superimposed onto a visual field and embodied interactions related to the auditory world of the audiobook in only the most accidental and arbitrary way. This overall depiction of audiobook listening is already implied by what we have discussed to this point in the chapter, but the following will lend it further support.

### 3. THE AUDIOBOOK HAS A SPEAKER

The audiobook is a performance, and both the performance and the performer contribute much to the listening experience. But what exactly is part

of the experience is not as clear as it might seem and depends on the listener as well as the performer. As Ihde has noted,

in a highly concentrated “narrow” focus I get certain sounds in the other’s speech but may find it almost impossible to note what was said; and contrarily in a “broader” focus, as in attending to what is being said, I may miss or barely be aware of the aspirated *s* which is characteristic of the other’s speaking style.<sup>19</sup>

Professional audiobook narrators are selected for their ability to perform with either a neutral or content-appropriate accent, in which the goal, presumably, is to minimize the likelihood that the listener’s focus will be drawn by these formal aspects of the performance rather than by the content which it is meant to represent. Authors who narrate their own works seem to be coached in the expected style.

The free recordings available online at LibriVox give us an opportunity to see whether and to what extent such formal variances influence our listening; LibriVox audiobooks are recorded by volunteers, and anybody with any kind of voice or accent may choose to record any public-domain work.<sup>20</sup> Although other listeners disagree, I have personally never found it distracting to hear P. G. Wodehouse performed with an American accent and see no reason why the Southern American accent should be limited to Southern genre pieces rather than, for example, a history of England or the Christian Bible.<sup>21</sup> Listening to Kirsten Ferreri’s recording of sections from *The Antichrist* on LibriVox was strange to me at first, but, in less than a minute, I noticed that I was no longer listening to a woman reading Nietzsche but was instead listening simply to Nietzsche. Alternately, listening to Gesine’s recording of Schopenhauer’s *Über die Weiber* did not produce the same transparency—that a woman had chosen to voice Schopenhauer’s most misogynist essay reminds the listener of the performer’s presence as a mute agent behind the voice, and the performer’s active personal silence in the verbatim production of words which she must clearly disavow makes the materiality of her voice productively obstructive rather than transparent.

Some narrators purposefully add distinctive vocal characteristics as an attempt to highlight and enhance content—“doing voices,” as for example with Jim Dale’s award-winning narration of J. K. Rowling’s Harry Potter books. In some cases the “voices” are simply a binding together and representation of the content already within the written word; in other cases the “voices” are more a creative superaddition supplying new content.

Altogether, the voice itself may be present in three primary ways: counterpunctual, consonant, and dissonant. It may be present as a representation and enhancement of content, much as a counterpoint supports a melody while creating new content in accordance with its form. It may be present although transparent, fading into the background of the listener’s

experience, allowing the listener to focus simply on the content, much as consonant notes accompany and harmonize with a melody, simply drawing out what is already in it. Or, it may be present in an obstructive form, whether in a purposeful and productive form, as in the example above of Gesine's Schopenhauer, or in an accidental and purely disruptive form, as might be the case with a title that has poor recording quality or a vocal characteristic which the listener is for some reason unable to allow to fade into the background. In either case, the voice is similar to dissonant notes which clash against and sour a melody—interesting and deepening the experience when this is the intended effect but less so when unintended or poorly done.

These three modes of interaction between form and content—counter-punctual, consonant, and dissonant—apply to physical and social contexts of listening as well. Listening to poetry on one's iPod may be given a strong counterpoint by walking through woods; may be simply consonant with driving, allowing the experience of the road to fade away into the background; and may be dissonant with the commuter crowd, either in a surreal and interesting manner or in a manner simply annoying and distracting.

#### **4. THE AUDIOBOOK IS STARTED, STOPPED, AND RESTARTED**

The listener must initiate the process of playback when listening to an audiobook, at which time the vocal performance follows its own predetermined schedule. The driver or the walker must plan ahead for the intended overlapping of experiences not only through the presence of audio files, with an appropriately integrated technological system of converting these files to sound waves, but also through the initiation of playback prior to the intended experiential convergence. The listener may experience frustration in starting playback in medias res, fumbling through CDs at a red light or trying to navigate MP3 directories while keeping pace.

If the driver does not specifically start and stop audiobooks but simply uses the powering on and off of the stereo as a stop and start button, the activity of driving is fully enclosed within playback: playback resumes upon starting the engine, and the driver must first attune to the audiobook, then turn to the road, and, once the driving process is complete, the playback continues until it too is halted. The walker, too, has experiences akin to these "driveway moments," in which the walker may continue on the exercise machine past the end of a planned amount of time or distance, waiting for an opportune and hard-to-anticipate moment of narrative resolution.

In the great majority of cases, the duration of playback is far longer than the duration of activities engaged in along with playback. The audiobook might reside in the driver's car or the walker's MP3 player for several weeks

and be played back over many commutes or exercise sessions. For this reason, along with those in the preceding paragraphs, it is appropriate to say that listening forms the context in which the driving and walking occur rather than the other way around. In Merleau-Ponty's visual terminology, the listening is the figure, and the driving or walking is the background. This conclusion may sound strange: we more often speak of listening to an audiobook while driving than of driving while listening to an audiobook. Still, this inversion is well supported by other common experiences: for example, that audiobooks may be most often listened to while on daily commutes or long road trips, situations in which little attention to the road is needed. It seems, further, that the very purpose of many who listen to audiobooks while at the gym—I suspect nearly everyone who does so—is precisely to move the process of exercise into the background in order to undergo it without so distinctly experiencing it. And thus the audiobook *forms* a context for physical and social experience rather than being experienced *within* a physical and social context.

Michael Bull has made observations in support of this point:

Listening frees up the eyes to observe and imagine, thus differing from the traditional reading of a book, in which the reader is visually engaged in the text. . . . The text becomes a continuous flow of sound on to which he adds a level of physicality in the act of imagination. The sound print of the book is imposed on the silence of the world around him.<sup>22</sup>

In the case considered here, the audiobook listener looks around a café, imagining that the people present are characters in the text. Clearly, there are many other ways that listening to an audiobook will condition physical and social experiences, depending upon the kind of listening involved and the kind of environment in which that listening takes place.

Ihde has discussed the orienting role of the background rhythms of nature and of the technosphere—the bird calls which signal the time of day, the hum of fluorescent bulbs which provide a context for paperwork.<sup>23</sup> When I plug my ears, I remove this context and its cues, and I replace them with auditory events occurring in a different environment and a different space. It may be that I am to some subconscious extent aware of the contours of that absent space to which I am listening in some dim shadow of the resonances made clear by composer Alvin Lucier's classic *I Am Sitting in a Room*—at least, I am aware that it is a space different from the one in which I am embodied and navigate, for it does not change its direction when I turn my head, nor does it fade as I walk, or change its resonance as I move from one room to another or go outside. In both time and space, the audiobook offers us an alternate and independent world that is literally experienced, without even considering the more metaphorical 'otherworldly' effects of narrative and imagination.

## 5. THE AUDIOBOOK FORMS A CONTEXT OF PHYSICAL AND SOCIAL EXPERIENCE

Some time ago, I walked through the grounds of the Bergianska trädgården in Stockholm while listening to a recording of Ralph Waldo Emerson's essay "Self-Reliance." Was one part of this experience supposed to enhance the other? Why is this "romantic" to say so? My memory of listening to Peter Kropotkin's *Conquest of Bread* associated with the woods around a particular stretch of Interstate 77 is no less distinct, nor is my memory of listening to William Sangster's *Umbrellas and Their History* while in a grocery store parking lot. Surely these moments cannot stand out in my mind due to any substantive connection between the physical experience and content of the audio context in which that experience occurred.

Less memorable is my experience of listening to David Hume's *Enquiry Concerning Human Understanding* while on the elliptical and the exercise cycle, not because there was less of a connection, or less of interest to remember from the book, but instead due to an excess of familiarity. I have met both Hume's text and the elliptical in different circumstances, and I can easily think of Hume without the gym coming to mind—but when I return to Emerson, the Bergianska trädgården comes along with it, as I have (as Hume would have put it) made them contiguous in my experience without diluting that contiguity through a multiplicity of other unrelated associations. "Custom," as Hume said, "is the great guide of human life."<sup>24</sup>

Certainly, as Bull has noted, audiobook listening changes the way we perceive our social environment.<sup>25</sup> To his example of the listener sitting in the café, we might add the listener who watches the bustle of crowds and the listener among the bustle of a crowd; the listener seated on public transit; the listener in the corner store or supermarket. As noted above, these experiences may be counterpunctual, consonant, or dissonant with the audiobook which provides their context. The difference, it seems to me, may depend on the interpretation of the text in which the listener engages as much as anything else. It is not just one thing to listen to Jane Austen on the subway.

Regardless of how the social environment is experienced by the listener, it is clear that the listener is in some kind of disconnection with the social environment, experiencing it within a context not available to others in that environment. If the context for being-with-others is supplied via earbuds, whatever experience of community the listener has must be a kind of false or imaginary community—a public space interpreted as a private and interior event. And yet the listener is also part of a real but nonlocal community, at minimum a community formed by the work itself: the author, the performer, the listener, and all the other unknown listeners. This community may be indefinite, where the listener might imagine how others reacted to, for example, a controversial event within a plot or claim within a nonfiction work. This community may be well defined, as when a listener

has picked up an audiobook on the basis of its bestseller status, was lent it by friends or family, or was recommended it by Oprah Winfrey.

And so here we see three kinds of community within the seemingly solitary and solitude-seeking act of listening to an audiobook in public: a real but nonlocal community which is formed around the aesthetic work; a local but imaginary community within the listener's privately contextualized experience of others; and a real and local but unexperienced community of audially unavailable mere presence-with others.

There are many further questions which a full phenomenology would address—in the preceding discussion I hope to have provided a solid start to such a phenomenology, along with a sketch of its contours. We have seen that the experience of an audiobook requires active binding of past, present, and future; requires us to construct a timeline separate from that of our physical interactive world; is constructed and bears authoritative expressions akin to written text; is personal and bears emotional attunement akin to spoken word; contains demanding and challenging silences; is borne by a voice which is counterpunctual, consonant, or dissonant; forms a context for physical and social experience; can be used to fade embodied experience into background; binds memory in ways very different from written text, due to the simultaneous experience of an arbitrarily related visual field; and produces unusual forms of isolation-in-community.

## NOTES

1. Remy C. Kwant, *From Phenomenology to Metaphysics* (Pittsburgh: Duquesne University Press, 1966), 158. Summary and representation of Maurice Merleau-Ponty's comments from "Interrogation and Intuition," in *The Visible and the Invisible*, trans. Claude Lefort (Evanston, Ill.: Northwestern University Press), 105–107.
2. Similar to many studies of phenomenology, this chapter uses single quotation marks in order to refer to something from within the phenomenological bracketing, or *epoché*. Single quotation marks allow the designation of a thing in its presence-to-experience without implying its absolute reality (as might be the case without any quotation marks) and without implying its illusory nature (as might be the case with double quotation marks). The single quotation marks allow us to signify a thing named while remaining uncommitted about whether this naming is a calling of a thing something which it is or a mere calling of a thing something which it is merely called.
3. Walter Ong, *Orality and Literacy: The Technologizing of the Word* (New York: Routledge, 1982), 31–32.
4. Maurice Merleau-Ponty, "The Film and the New Psychology," in *Sense and Non-Sense*, trans. Hubert L. Dreyfus and Patricia Allen Dreyfus (Evanston, Ill.: Northwestern University Press, 1964), 54.
5. Merleau-Ponty, "The Film and the New Psychology," 49.
6. Don Ihde, *Listening and Voice* (Athens: Ohio University Press), 89.
7. John B. Brough, "The Phenomenology of Internal Time-Consciousness," in *Husserl: Shorter Works*, ed. Peter McCormick and Frederick A. Elliston (Notre Dame, Ind.: Notre Dame University Press, 1981), 272.



8. Edmund Husserl, "The Lectures on Internal Time-Consciousness from the Year 1905," in *Husserl: Shorter Works*, ed. Peter McCormick and Frederick A. Elliston (Notre Dame, Ind.: Notre Dame University Press, 1981), 283.
9. Maurice Merleau-Ponty, *Phenomenology of Perception*, trans. Colin Smith (New York: Routledge, 1998), 418–419. The translation of "*Abschattungen*" is my own.
10. Thomas Clifton, *Music as Heard: A Study in Applied Phenomenology* (New Haven: Yale University Press, 1983), 62.
11. John Stuart Mill, *Utilitarianism*, 3d ed. (London: Longmans, Green, Reader, and Dyer, 1867), 17.
12. Do you read a footnote immediately, in the middle of a sentence, or afterwards? Where do you go back to when you return to the sentence? Do you have to start back at its beginning? Are you supposed to reread the sentence with the footnote retained as a kind of semi-literal "subtext"? How different this is from the spoken tangential remark!
13. Ihde, *Listening and Voice*, 154.
14. Martin Heidegger, *Being and Time*, trans. John Macquarrie and Edward Robinson (New York: Harper & Row, 1962), H. 134.
15. Marshall McLuhan, *Understanding Media* (New York: McGraw-Hill, 1964), 82.
16. Edmund Husserl, "The Lectures on Internal Time-Consciousness," 281.
17. See Heidegger, *Being and Time*, H. 243–4.
18. Nick Piombino, "The Aural Ellipsis and the Nature of Listening," in *Close Listening: Poetry and the Performed Word*, ed. Charles Bernstein (New York: Oxford University Press, 1998), 70.
19. Ihde, *Listening and Voice*, 89.
20. LibriVox, <https://librivox.org>. For more on LibriVox, see Michael Hancher's chapter in this volume.
21. It is worth noting that some listeners attempting to use audiobooks for the purpose of language acquisition have found special value in the way that LibriVox recordings represent a variety of different forms of spoken English, and make use of LibriVox recordings specifically to gain access to accents prevalent in spoken English but not favored in commercially produced speech.
22. Michael Bull, *Sound Moves: iPod Culture and Urban Experience* (New York: Routledge, 2007), 40–41.
23. Ihde, *Listening and Voice*, 86.
24. David Hume, "An Enquiry Concerning Human Understanding," from *Enquiries Concerning the Human Understanding and Concerning the Principles of Morals*, ed. L. A. Selby-Bigge, 2d ed. (Oxford: Clarendon Press, 1902), §36, 44.
25. Bull, *Sound Moves*, 40–1.
26. Here, I have in mind Ruud Kaulingfreks and Samantha Warren's discussion of community and solitude in iPod listening, where the first kind of community corresponds to their discussion of Hans-Georg Gadamer's *The Relevance of the Beautiful and Other Essays* (Cambridge: Cambridge University Press, 1986), and the third kind corresponds to their discussion of Jean-Luc Nancy's *The Inoperative Community* (Minneapolis: University of Minnesota Press, 1991). See "Mobile Clubbing: iPod, Solitude, and Community," in *iPod and Philosophy*, ed. D. E. Wittkower (Chicago: Open Court, 2008), 167–179.