
NOT JUST ONE,
NOT JUST NOW
Relational Voices in Time

MATTHEW RAHAIM

FEW philosophical themes evoke as much wonder as voice. There has been a long tradition of phenomenological reflection on vocal utterance: from Maurice Merleau-Ponty's evocative image of sonorous speech as "singing the world" (Merleau-Ponty [1945] 2012, 193) to Emmanuel Levinas's open-ended "traumatism of astonishment" that arises in face-to-face conversation (Levinas [1961] 1979, 73); from Don Ihde's living, breathing cosmic vocality, in which we are surrounded by the voices not only of people but also of sliding mugs, rolling dice, and popping firecrackers (Ihde [1976] 2007, 67), to Jean-Luc Nancy's rapturous reflections on "Mmmmmmm," the syllable that "resounds previous to the voice, inside the throat, scarcely grazing the lips from the back of the mouth, without any movement of the tongue, just a column of air pushed from the chest in the sonorous cavity, the cave of the mouth that does not speak" (Nancy [2002] 2007, 24–25).

But one particularly persuasive treatise on the phenomenology of voice seems to drown out all the others: Jacques Derrida's analysis of vocal presence in *Voice and Phenomenon* ([1967] 2011).¹ This work, and the critique of logocentrism that it inaugurated, is often construed as a damning rebuke, revealing both phenomenology and vocal philosophy to be "outmoded and backward," childishly attached to the voice's false metaphysical promises of presence, immediacy, and *logos* (Kane 2015, 672). At the very least, it shows Derrida to be a competent phenomenologist in his own right and an exceedingly subtle reader of Husserl in particular. *Voice and Phenomenon* focuses on one particular case of voicing that, Derrida argues, is fundamental to any descriptive phenomenology: a subject speaking to itself, cut off, in principle, from relations with others. Though this self-absorbed speech seems to offer an immediate, immaculate *expression* of inner meaning, Derrida demonstrates on phenomenological grounds that even the apparently pure voice of soliloquy has its own contingencies, unavoidably entangled in temporality, alterity, and interpretation. And if even speaking to ourselves is so messy,

what of speech and song out there in the world—so often construed as originary, pure, and authentic in comparison to writing?

There is much for scholars of voice to learn from Derrida's critique. Husserl is not alone in figuring the inner voice as an unmediated source of evidence.² A surprising amount of discourse about "the voice" implicitly frames vocality as an expression of a sovereign individual or a monolithic collectivity, sounding out a pristine self-identity, insulated from the complex interplay of social relations. This expressivist figuration guides political discourse too. We speak casually—too casually, perhaps—of the unified "voice of the people," the unique "voice of a nation," of politicians who "give voice" to a putatively unanimous constituency, as though unblemished self-identical unison were the normative prior condition of any politics.

But we have much to teach as well. Public singing and speaking, chatting and harmonizing with others, vocal uproar, protest, and negotiation, so central to any ethnomusicological phenomenology of voice, appear in the Derridean critique only as exceptions to the Husserlian rule—or, as Derrida would have it, as manifestations of a single metaphysical "impurity": *absence* ([1967] 2011, 73). Ethnomusicologists, in contrast, have long afforded primacy to the social life of song, amplifying the life of voices bouncing relationally about, feeding back, plying their various trades among *others*: declaiming, seducing, transgressing, drawing together consociates. Michelle Kisiuk's classic work on Pygmy song focuses on vocal relationships rather than mere expressions of a pregiven social order, opening up a performative, temporally dynamic "micropolitics" (Kisiuk 2000, 26). Steven Feld's paradigmatic case of *groove* (by which sound structure *is* social structure) is a particular Kaluli practice of collective song (*dulugu ganalan*, or "lift-up-over sounding") that sounds out, negotiates, and cultivates (rather than merely reflects) a participatory, anarchic, egalitarian way of being (Feld 1984). The past few decades have seen an acceleration of focused ethnographic work on voice: on the ethical stakes of vocal comportment (e.g., Desai-Stevens 2017; Eidsheim 2015), the formation of vocal subjects (e.g., Weidman 2011; Harkness 2013), and the politics of public voicing (e.g., Kunreuther 2014; Schultz 2012). On the whole, this literature foregrounds the ongoing dynamism and plurality of vocal action, rather than construing the voice as an instrument of immediate expression.

This essay attempts to bring this ethnographic literature into conversation with Derrida's critique, suggesting ways in which an ethnographically grounded phenomenology of voice might highlight precisely the kinds of social relations that Husserlian soliloquy excludes. It also points to some of the politically consequential ways in which regnant social-theoretic construals of voicing presume a notionally atemporal, eternally present, perfectly expressive soliloquy. Our trajectory begins with the interwoven figurations of voice, temporality, and melody that underlie the Husserlian-Derridean thread and which seem to yield a figuration of voice that is utterly irrelational, cut off from social relations. We then turn to the temporality of vocal co-presence—the intersecting arcs of call, response, interruption, and affirmation. By the end, we will return to the temporality of relational voicing, its ontological possibilities, and its contingent forms of immediacy.

THE VOICE OF PHENOMENOLOGY

In what sense does a voice seem to offer immediate presence? Simon Frith suggests one way: since “we have bodies too, throats and stomachs and lungs . . . the voice seems so directly expressive [because] it doesn’t take thought to know how that vocal noise was made” (1998, 192). In another obvious sense, the presence of a speaker can be a compelling illusion—for example, in the famously fraudulent vocal performances of Orson Welles’s *War of the Worlds*, which was broadcast on CBS radio in 1938. The actor playing a news anchorman purports to offer an eyewitness report of a spaceship landing in New Jersey, narrating everything that he sees, at the very moment he is seeing it: “I can see peering out of that black hole two luminous disks. Are they eyes? It might be a face. . . . There, I can see the thing’s body. It’s large, large as a bear and it glistens like wet leather.” So convincing was this first-person present narration of fictional *seeing* that some listeners seem to have taken it to be a real news broadcast. Even listening to this broadcast a century later, knowing full well that the report is fictional, we are still offered a vivid picture of an event unfolding *now*. Even a high-resolution photograph of a person—flat, silent, and still—lacks the absorbing sense of immediacy offered by a voice narrating in first-person present tense.

But there is a subtler sense in which a voice may seem to offer itself up as presence itself, and this is the theme that Derrida takes up in *Voice and Phenomenon*, a dense extended essay that provides the phenomenological grounding for Derrida’s famous critique of the *metaphysics of presence*. Derrida is broadly concerned with the tendency to assign metaphysical priority to the living, sounding, immediate presence of speech rather than the spooky ambiguities, slippages, and silent absences of writing. Derrida’s oft-cited historical claim that “Western” metaphysics has, since Plato, privileged the voice over writing turns out to be cherry-picked and incomplete at best (Cavarero 2005, 227–234; Dolar 2006, 43–52). The enduring value of Derrida’s argument instead lies in his phenomenological analysis. As Kurt Gödel’s *Formally Undecidable Propositions* did for Whitehead and Russell’s mathematical logic, Derrida’s *Voice and Phenomenon* probes the very fabric of Husserlian phenomenology using its own methods—and discovers a seam. In order to find it, as Derrida teaches us, we must listen (Derrida [1967] 2011, 74).

Derrida argues that phenomenology is founded upon a very particular kind of speech: *soliloquy*, or speaking to oneself. He begins from Husserl’s early foray into semiotics in his *Logical Investigations*, which seems to grant soliloquy a kind of semiotic immediacy that colloquy (interpersonal speech) cannot have (Husserl [1900] 2001, 218–219). Husserl claims that while colloquy *indicates* a meaning through signs to someone else, soliloquy is radically different: it *expresses* a prior meaning that the speaker has already grasped in its totality. While we grasp the voice of an interlocutor partially and indirectly (following along the trail of the words, always a bit late), soliloquy seems to flash into being precisely at the same time as its meaning (“in the twinkling of an eye”). In Husserl’s early semiotics, soliloquy is perfectly expressive: a meaning-to-say

is expressed immediately in its entirety, without distortion, without stopping to work through what individual words indicate. Derrida argues that this figuration of apparently direct vocal expressiveness, which seems to promise immediacy, presence, and unmodified access to *logos*, is the characteristic condition of supposedly pure phenomenology. He highlights the peculiar lilt of phenomenological prose, which generally proceeds in the first person, in present tense, as, for example: “I see in pure reflection that this die is given continuously as an objective unity” (Husserl [1913] 1982, 39). Like the virtual journalistic reportage of *War of the Worlds*, the validity of this voice does not hinge on its being factually true but rather on its seemingly immediate access to its own experience. Of course, the phenomenological philosophy we encounter written on the page is inevitably indicative, rather than expressive, in Husserl’s scheme: it re-presents the now-absent originary moment. Yet if it is to be a pure phenomenological account, its authority would seem to depend on an originary moment of soliloquy, in which the self speaks to itself in a field of unmediated self-presence, in which “the acts in question are themselves experienced by us at that very moment” (Husserl [1900] 2001, 191).

Temporality, then, is at the heart of this construal of phenomenological voice. As Derrida points out, though, we encounter a paradox. By Husserl’s own later analysis of temporality, any lived present moment depends on retention of the just-past and protention of the just-about-to-come. In other words, even the presence of one’s own voice requires the re-presentation of “absent” moments. To understand this temporal paradox, it turns out that once again, we must listen—in this case, to melody.

TWO MELODIC TEMPORALITIES

Melody has served as a model for many classic accounts of temporality, from Hume to Brentano, Bergson, Husserl, and beyond (Hume [1738] 1960, 36; Brentano [1874] 1995, 217; Bergson [1889] 2001, 100; Husserl [1928] 1991). Despite their vast mutual differences, nearly every such account tends to share a single model of melody: an evenly spaced sequence of discrete notes, plunked out on a piano, one by one.³ Each tone has a clear beginning and end; each is the same length; none has any rhythmic or gestural priority over others. For philosophers, this paradigm is heuristically convenient, offering a comforting stability in an otherwise dynamic field of perception, providing (it seems) a simple set of building blocks for temporality to operate upon. For practicing musicians, however, such an account of melody can only appear naïve. Its peculiar temporal affordances (lacking in what is casually called “musical time”) has led to a crucial paradox in the phenomenology of temporality. Call this model of melody (as a string of atemporal, notionally interchangeable objects) a *punctual* melodic paradigm; we will return to it and its implications for temporality later.

First, we turn to a rather different melodic paradigm—a *vocal-gestural* paradigm, in which any melodic utterance is actively performed over time. The voice projects rather different temporal possibilities than the ideal-typical piano assumed by most

philosophers.⁴ What is it to speak a word, to sing a word, to hear a word? It differs from the instantaneous *logos* of Husserlian soliloquy and from the digital plunking of discrete notes in at least one important way: it takes time. No one beeps out a note with the perfect acoustical precision of a tuning fork. Even the most tuneful singer overshoots minutely and corrects, or hovers around their destination with vibrato, or scoops or glides from tone to tone. Speech too is gestural. Consider the word “time”—or, better, say it. The plosive /t/ requires a buildup of air pressure for fifty milliseconds or so, and then releases it. After a barely discernible moment of hissing air, the vocal cords gradually engage on the vowel, a diphthong, moving from /a/ to /i:/. No sooner does the /a/ form than the back of the tongue begins pulling up to form /i:/; /m/ is enunciated by a gradual closing of the lips and a brief hum. Typically, at the end of a sentence, the word is finished off with a slight opening of the lips at the end, articulating a gentle, subtle “muh.” We sing it too; it may rise (“Is it *time*?”) or fall (“It’s *time*.”) (see Figure 13.1), or trace subtle trajectories of irony, taunting, or coaxing. In the figure, the successive phonemes are shown in noteheads for convenience, but the real action is shown in the spectrograph

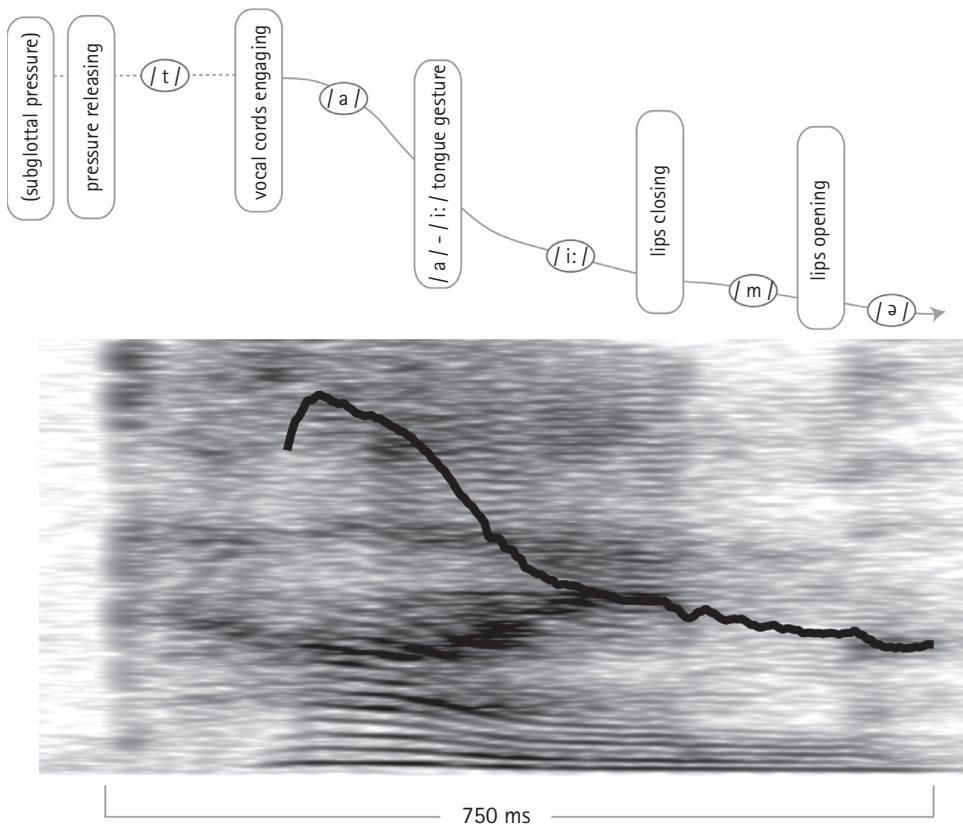


FIGURE 13.1: Saying the word “time.” Here, the thick black line is the pitch trace; the background is a spectrogram showing the gradual shifts in overtones from vowel to vowel.

(the comblike striations that show timbre) and the pitch trace (the thick black curve): both are continuously shifting, articulating gestural motion rather than points.

The melodic shapes that allow us to distinguish a question from a solemn oath from a mocking impersonation unfold in expansive curves. These utterances become meaningful only through a temporal sense of where the utterance is going. These are constantly in motion and emerge with duration, not as instantaneous, discrete points. An infinitesimal moment could never suffice for this coordinated muscular action, nor could a millisecond, nor even ten. No solitary point on a horizontal axis can give even a single syllable; every real utterance has its own inherent temporal thickness.

Utterance furthermore maps out a temporal itinerary for itself. The thrill offered by minced oaths such as *shut the f. . . ront door* rests on the almost irresistible anticipation (*protention*) of the profanity heralded by /f/. The timing of a punchline relies on a magical play with an audience's expectations (Berger 2010, 48). Elaborate melodic action plays on just these expectations. When we hear a cohesive melody, we are already carrying out a remarkable act of temporal constitution: *retaining* the residual presence of the melodic arc and *protending* the arc to come. Thus, melodic apprehension—like any temporality—emerges in a presence that is inescapably “thick” (constituted by disciplined retentions and protentions), rather than “thin” (locked in a single, infinitesimally brief moment of perfect immediacy). These presences may come in various extents (ranging from a few hundred milliseconds to several seconds) and may carry various affective valences. We may feel curiosity, yearning, or dread for what is to come; we may delight in the elegant completion of a rhyming couplet, a modulation, or a *tihāi*. In any case, the attentional disciplines that lead to particular forms of protention and retention are not mechanically triggered by the sonic signal but require enculturation and training.

For example, Figure 13.2 shows a pair of phrases (A and B) sung by Amir Khansahab, one of the best-known Hindustani vocalists of the twentieth century, singing in Rāg Rāgeshrī. Phrases like this are *extemporaneously* generated—which is to say that Hindustani vocalists come up with them “at the moment.” As should already be clear, this moment could never be infinitely thin; it already builds on the retained past and projects a protended future.

As with most extemporaneous utterances, spoken or sung, the hands are working alongside the voice (McNeill 1992; Kendon 2004; Streeck 2009). As in sentences, the hands trace breathing, neumatic arcs of rise and fall, antecedent and consequent, tension and release. In this case, they are playing with the temporal expectation of return to the tonic from an excursion. Phrase A in Figure 13.2 moves from tension to relaxation (see Figure 13.3). As Khansahab's right hand grips and gathers a virtual substance, the wrist and finger joints move from flexion (with muscle tension pulling toward the palm and little finger, with a loose grip) to neutral relaxation (wrist in line with the forearm, open handed).

Meanwhile, his voice moves along an oft-traveled path to the tonic, a modally specific cadence already burned into his listeners' sensibilities: *ga ma re sa*, {3 4 2 1}.⁵ Or so we expect. Phrase A is different; he begins on the *ga ma re sa* path but slips *below sa* to touch *ni* {7_b} as his hand extends back the *other way*. He then overshoots again, this time above the tonic, flexing his wrist accordingly, before returning to *sa* {1} and relaxing his wrist.

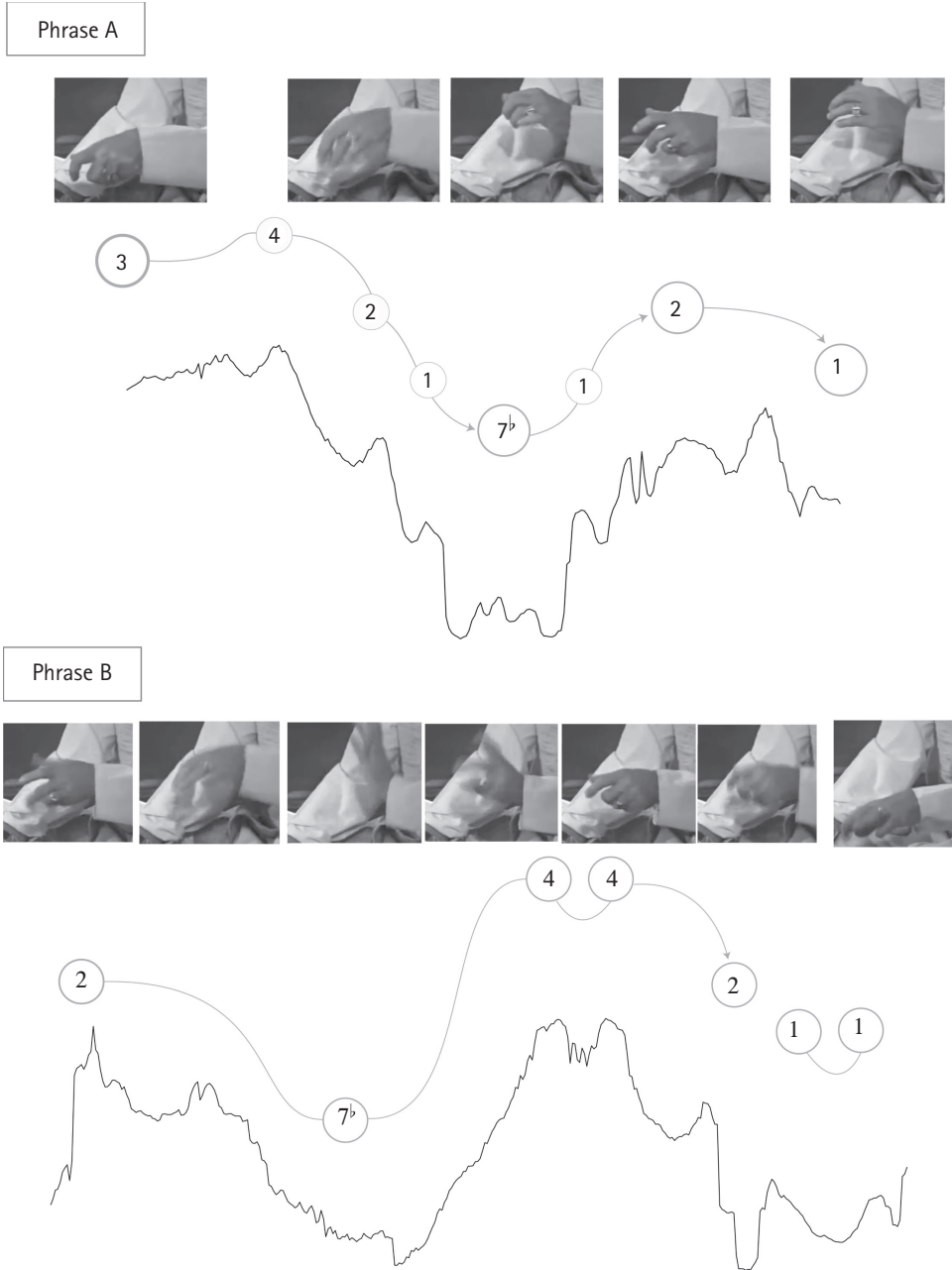


FIGURE 13.2: Phrases A and B. Amir Khan, Rāg Rāgeshrī. Arabic numerals indicate scale degrees.

But even here, his arm has not fully relaxed; his hand is raised, cupped, and ready for phrase B to complete it. Phrase B too delays the resolution to *sa* with a dramatic *re ni ma* {2 7^b 4} swoop before coming to repose on the tonic.

To feel the melodic ebb and flow of these phrases, it is not enough to ride their pitch contours like a rollercoaster, locked utterly “in the moment,” trapped in the flux of pure

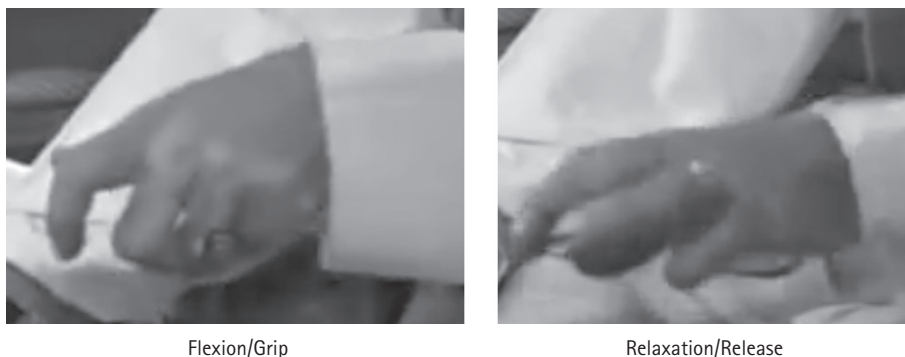


FIGURE 13.3: Flexion and relaxation. Amir Khan, Rāg Rāgeshri.

awareness. As with the vocal-gestural utterance of a word, the melodic action in these phrases is all in playing with a listener’s melodic expectation: their retention of the trajectory just past, and their protention of the resolution about to come.

Though often depicted as purely cognitive, protention is also a physical act. A gestural handshape forms protentively *before* it is deployed; as Merleau-Ponty observes, “from its very beginnings, the grasping movement is magically complete; it only gets under way by anticipating its goal” (Merleau-Ponty [1945] 2012, 106). Connoisseurs of raga music often mark culminations of metric cycles by raising and dropping their hand: the fulfilment of the cycle is felt in advance, and the hand is raised and prepared for the drop. As a goalie raises her hands before the instant of contact with the ball, protention often happens physically, publicly, even socially, as groups of connoisseurs “tune in” to the temporality of metric cycles, feeling the culmination together, moving as one. Breath too protends the melody to come. A singer opens her mouth and takes a breath in advance of singing a phrase. And once singing, she is always faced with a point on the temporal horizon when the next breath must come. This is no less true of instrumentalists. A guitarist playing in a “lyrical” style, a sarangi player evoking a wail or a sigh, and a sitarist playing in *gayaki ang* (a “singing way”) all project an eminently gestural melodic logic. The hand of a pianist rises in preparation for a downbeat, falls, roots the thumb on D in preparation for an ascending tetrachord from E to A, and traces arcs across the face of the keyboard (Sudnow 1978). The temporality of the musicking body is the temporality of preparatory breaths, manual retention, dramatic satisfaction, or deferral of anticipations.

PARENTHESES: EMBODIMENT, STILLNESS, AND DISCLOSURE

To summarize this rich kinetic detail by saying that music is “embodied” is a decent provisional gloss, and a great deal of insightful literature has recently come under

this rubric. However, the excessive polysemy of “embodiment” can lead to various misreadings, and it’s worth pausing to comment on these.

A reliance on the term “embodiment” may suggest a pre-given anatomy shared by all singers—or worse, that each body begins and ends at the individual. But musicking bodies are plural, widely various, and socially enculturated (Rahaim 2012; Weidman 2012). And they are productive, in turn, of various vocal ontologies, according to their techniques (discussed later). “Embodiment” may also suggest a metaphysics in which some prior “disembodied” music is optionally and occasionally instantiated (“embodied”) by some body. But having a body is a condition of possibility for the disclosure of a musical world in the first place: whether dancing, playing, singing, or listening, we are engaged in bodily acts. Perhaps most misleadingly, speaking of embodiment may suggest that some activities are embodied and some are not, as though sitting very still while singing is “disembodied” in contrast to singing while moving. But sitting still for an extended period of time is also a cultivated technique of the body, far more difficult to instill in children, for example, than clapping or dancing. The retrained stillness of Amália Rodrigues as she sings fado, for example, contrasts with the usual expressive gestures one would find in a spoken lament, complaint, or exclamation; this is an eminently disciplined musicking body, not its ghostly disappearance. Extended *padmasana* (i.e., yogic “lotus posture”) affords meditative practices that disclose an inner landscape that is inaccessible while navigating a busy street; extended *sitzfleisch* (i.e., a still and quiet sitting posture) in the concert hall can disclose the forms and objects on which a distanced aesthetics can operate. Bodily techniques of stillness are world-disclosing practices no less than bodily techniques of overt motion.

Among the things that still contemplation can disclose is a world of melodic objects—sequences of notes, brightly salient against a background—that seem to outshine the kinetics of motion. The notes transcribed in Figure 13.2, like the phonemes in Figure 13.1, are actively discerned by an enculturated listener, not given in the sonic signal. Nowhere in the undulations of vocal pitch do we find a flat, steady line that simply corresponds objectively to a “note.” The disclosure of notes requires repeated listening and study, sometimes with eyes closed, speeding up and slowing down a recording, testing my provisional notations against listening, until I am acceptably content with the result. A trained ear can discern them, to be sure, and this discernment requires considerable work. (This is why “ear training” requires years of disciplined practice even for skilled singers.) Not all—or even most—singers have this skill; few, if any, can achieve good melodic results by willfully piecing melodies together note by note.⁶

PUNCTUALITY, NOTATION, AND PRESENCE

Note-by-note is the logic of the punctual melodic paradigm outlined earlier. A punctual account of a melody does not work from a particular performance but from a sequence of notes, fixed and given like a string of beads. To be sure, these punctual accounts are directed toward uncovering the thickness of the retentional-potential

present (Husserl [1928] 1991, 11). To hear a melodic sequence, for example, {12323, 23434, 34545 . . . }, requires the retention of past phrases; harmonic tension derives its effect precisely from the protention of a resolution to come. And yet, oddly enough, the punctual paradigm can only offer flat, homogeneous sequences, without rhythmic impulse, speed, goal tones, energetic flow, or improvisational spontaneity. Even when Husserl's account stresses the fading away of tonal impressions over time, the fading is "incessant," each tone notionally equivalent and interchangeable in its linear decay, independent of phrasing, metric cycle, or breath, "until it disappears" ([1928] 1991, 32). In other words, such a temporality is grounded in tones that are, themselves, remarkably devoid of action, assuming, as Indian music theorist B. C. Deva puts it, "a spurious constancy of tones and tonal relations" (1959, 175). In such a conception, tones sit perfectly still, awaiting our perceptual uptake into a temporal schema. They occur in a definite, predictable order, without any need for a trained listener to make sense of them.⁷ While conventional staff notation usually includes at least some indications of phrasing, the discrete tones that underlie the melodic accounts of well-intentioned philosophers typically do not. They have the same temporal ambiguity as a short series of photographs. As Hindustani musician and theorist Sakhawat Hussain Khan once vividly put it, sequential notation is a sort of vocal photography⁸ that freezes the dynamic flux of singing frame by frame, so that a song can be learned instant by instant, note by note. Indeed, the usefulness of sequential notation is precisely that it may be studied and learned outside of the flow of melody. It does this by reconstituting the continuity of melodic motion as a series of successive snapshots of indeterminate spans of time. As photography obscures motion and generates an ambiguous sense of an infinitely still "instant" otherwise inaccessible to experience, so too does tone-based notation generate a punctual melodic paradigm.

We now begin to see the outlines of the paradox produced by the punctual melodic paradigm. Even if retention and protention are necessary to disclose a melody from a sequence of tones, for Husserl, "particular acts correspond to the individual tones" (Husserl [1928] 1991, 141). These acts (of attention, focus, and discernment) thus become tied to the curious entity that Husserl calls the *primal impression*—the "consciousness of the tone-now" that, along with retention and protention, completes his triadic model of temporal structure ([1928] 1991, 31). The primal impression poses extraordinarily difficult philosophical problems, as both Husserl and Derrida acknowledge; here, "words fail us" (Derrida [1967] 2011, 72n1). It is not clear from Husserl's text, for example, whether each primal impression is (1) a punctual, immediate experiential snapshot of the now-point that passes and instantaneously transforms into a retention, to be replaced in turn by another primal impression (and another, and another), or (2) a durative ("thick") moment, constituted temporally by a subject, always already bundled together with retention and protention (Zahavi 1999, 65). The term "primal impression" does read at times as though it occurs as a point (like a notional note-punctum), or as Derrida puts it with great self-assurance, "the identity of experience instantaneously present to itself" (Derrida [1967] 2011, 60).⁹ When Husserl describes the constitution of motion (a far more intuitively musical situation than the tickertape "melody" described earlier), he

has recourse to a “nucleus” in the midst of this continual flow, a “head attached to the comet’s tail of retentions,” and a “grasping-as-now that takes place moment by moment” ([1928] 1991, 32). As in the algebraic analysis of curves, we appear to have a play between the apprehension of a continuous span of motion (“during the time,” “the comet’s tail”) and numerous discrete points by which the curve becomes knowable in primal impression after primal impression (i.e., a “grasping-as-now,” “moment-by-moment”). There have been various readings of this ambiguity. On Derrida’s reading, Husserl’s primal impression is meant to be infinitely thin and nailed to the now-point, untouched by retention and protention. It is the temporal condition of soliloquy’s self-presence: pure “auto-affection,” which is “affected there by nothing other than by itself” (Derrida [1967] 2011, 73). And yet, were this actually the case, Derrida points out that the “itself” by which each primal impression is successively “affected” (i.e., replaced in turn), and by virtue of which it is a “now,” is yet *another* successive primal impression. This supposedly pure self-presence, then, is shown to be always already founded on difference with something else. In these terms, Derrida is able to persuasively debunk the implication of immediate self-presence in Husserl’s earlier account of inner speech: “the theme of a pure interiority of speech, or of ‘hearing-oneself-speak,’ is radically contradicted by ‘time’ itself [i.e., by temporality]” ([1967] 2011, 74).

But here is the paradox: in a punctual melodic paradigm, these two readings of the primal impression are indistinguishable, as tonal durations are indistinguishable from tonal instants. Since a sequence of internally homogeneous tones yields no difference, no change, over the span of each of its atomic tones, it could only offer a sequence of snapshots of pure “nows.” Melody (which serves for all of these philosophers as a metonym for temporality in general) can only emerge sequentially, point by point. Each “tone” could, in principle, last for a minute or a millisecond; it could be sustained right up until the following tone or appear to be infinitesimally short. Within each point, temporality seems to disappear, deprived of any difference that would yield a sense of motion, progress, or change. It is indeed possible to approach this form of extraordinary temporality as a sort of psychedelic limit case—as, for example, in Steve Reich’s “Pulse,” Terry Riley’s “Desert of Ice,” or other forms of pointillistic minimalism that are remarkable, in part, for their uncanny distance from gestural utterance. As we have seen, though, this is not an accurate model of how song, speech, or even words actually unfold in performance. When the voice does come to rest on a tone (what is called a *nyāsa svāra* in Hindustani music), it shows up as a distinctive noema, a special object, that stands out against a background of continuous motion (Rahaim 2012, 95). In a gestural melodic paradigm, by contrast, there could be no “thin” now-point, no primal impression independent of retention and protention.

It is hard to sustain Derrida’s certainty about Husserl’s intent, especially since Husserl himself was never satisfied with his solution (Zahavi 1999, 69–75). Indeed, his work starts from an explicit rejection of Brentano’s theretofore regnant distinction between direct “perceptions” of the infinitesimal present and the indirect “presentiating acts” of retention and protention—a distinction he rejects precisely because it cannot account for the immediate perception of temporally unfolding events such as melody.

Indeed, the entirety of his work on temporality may be read as an insistence on a “thick” temporal presence. And yet, Husserl’s famous term for the moment of the primary impression, “*im selben Augenblick*” (“in the glance of an eye”), seems to ride at the edge of “thick” and “thin” presences (durative yet almost imperceptibly brief).¹⁰ Derrida conveniently renders *Augenblick* as “blink” (i.e., the closing of the eyelid, like the shutter of a camera) and then goes on to dedicate the fifth chapter of *Voice and Phenomenon* to the implications of this figment of his own creative mistranslation.¹¹ His aphoristic response comes in two phrases. The first phrase, “there is a duration to the blink,” correctly emphasizes the temporal thickness of protention and retention. The second phrase, “the duration closes the eye,” however, implies that protention and retention introduce an antagonistic *absence* ([1967] 2011, 56). Absence is Derrida’s favored all-purpose metaphysical substance; it seems to encompass retention, memory, protention, anticipation, and, in the end, the thick present itself. Indeed, Derrida extends the metaphysical ambit of presence/absence so far that, in his later work, it seems perfectly natural for him to conflate, paradoxically, the always-incomplete gestural unfolding of neumes (vocal gestures) with Rousseau’s fantasy of a “time reduced to presence” ([1967] 1997, 249).

The rhetorical force of *Voice and Phenomenon* largely rests on the imputation of a foundational, puritanical zeal to Husserl’s work, as though phenomenology depended on sealing off the inner voice against what Derrida calls “contamination” by absence ([1967] 2011, 19). It is not clear that this was ever Husserl’s position. More importantly, as Adriana Cavarero so compellingly demonstrates, Derrida’s tightly focused analysis of soliloquy still “imprisons [voice] in the very metaphysical box that it was meant to disturb” (2005, 215). Even Derridean absence, after all, does not quite offer relationality.

THE POLITICS OF VOCAL IRRELATIONALITY

For ethnographers of voice, the resounding import of *Voice and Phenomenon* lies not so much in its successful indictment of phenomenology or of voice studies, but in its caution against a commonsense metaphysics that figures The Voice as solitary, immediate, and utterly free from relations. Such a “voice” is a ghostly caricature, not an ethnographic account of singing, arguing, or greeting. Indeed, it is inadequate even for the silent soliloquy of inner speech; it assumes that each subject has a single, eternally present, inner self who speaks and listens with absolute certainty. But numerous accounts of inner voice attest to polyvocality—plural inner voices that seem to interact, that take various stances, and that are often at odds with each other (Alderson-Day and Fernyhough 2015; Bakhtin 1981; Vygotsky [1934] 1987). One of the most striking empirical confirmations of polyvocality is found in Tanya Luhrmann’s ethnographic work with evangelical Christians who train themselves to discern the voice of God from the voices of other various inner speaking selves (Luhrmann 2012). Luhrmann’s work not

only affirms the fact of polyvocal inner worlds but also emphasizes the need for *training* to discern a single voice among them. Husserlian soliloquy thus serves as a theoretical limit case, like a perfect vacuum or a tree falling alone in the forest. It forces us to confront the eeriness of a voice stripped of all forms of relation: solitary, purely expressive, atemporal—in other words, irrelational. As Cavarero points out, commitment to such a figuration comes at a very high ethical and political price: “the elimination of the other [and] of others” (2005, 46).

This ghostly, irrelational picture of voice still haunts us. It can be found, for example, in figurations in which “voice” is understood to be the direct expression of a sovereign subjectivity. The unambiguous ascription of voices to individual subjects dovetails conveniently with a liberal ethics in which it is morally incumbent upon each individual to find and express their unique voice. Indeed, as James Faubion has pointed out, this is one of the grand thematical demands of our age (2011, 258).

This solipsistic ethics of voice is far more serious, however, in the service of collectivist nationalisms, where the putative sovereignty of individual expression is projected onto large, notionally homogeneous groups of people, claiming a single voice for a sovereign nation-state, culture, or society. This “groupism” (Brubaker 2002) has been strangely persistent in ethnomusicology. It is common to encounter majoritarian nationalist ascriptions of voices to putatively homogenous nations, as with Umm Kulthum in Egypt or Lata Mangeshkar in India (Danielson 1998; Srivastava 2004; Lomax 2003, 267). Even more common are claims about the “voice of the people.” Like invocations of putative “national conversations,” these claims point vaguely toward thousands or millions of people. But a “voice of the people” does not have conversations. It may *refer* to an alien group, but it does not ordinarily *address* them, listen, or respond. In such an expressivist frame, any actual social relations that might obtain *between* voices is blotted out by a prior, encompassing, irrelational oneness. When a voicing group is construed as a single unanimous totality, it serves, social-theoretically, as a sovereign, self-identical individual in its own right, with perfect unmediated access to its own inner workings, and a single expressive voice. The result is structurally identical to Husserlian soliloquy: the self expresses the self to the self out of time, out of history, untouched by alterity, with no peer or interlocutor with which it can be in ensemble. At the turn of the twenty-first century, such vocal nationalisms might have seemed quaint, but the worldwide resurgence of ethnic nationalisms has lately infused them with new blood.

To construe the voice as immediately expressive of a nation-state, then, is to naturalize a collectivity that is, itself, peculiarly irrelational and devoid of sociality. “The voice of the people” isn’t always congruent with the voice of the state; it has, at times, served to assert a contingent, tactical popular solidarity against oppression, domination, and colonial occupation. But this collectivist-expressivist vocal figuration is the quintessential “master’s tool” (Lorde 2020, 102). In its increasingly common majoritarian statist figurations, insisting on timeless self-identity and the direct expressiveness of a collective voice is to deny the very possibility of dissent or diversity.

VOCAL RELATIONALITY

As ethnomusicologists, for better or for worse, we mostly work among chaotic reverberations, amid mediated feedback, and in dynamic relationship with other people: in town squares, in living rooms, at political demonstrations, in nightclubs. And these forms of sociality, as the essays in this volume attest again and again, are enriched, rather than undermined, by the situated practice of ethnographically oriented phenomenology, attuned to social relations rather than to soliloquy. This performed sociality is a rather different matter than the common casual recourse to “social” collectivities (discussed earlier in relation to national “voices”), which begin by assuming a prior, univocal, fixed society,¹² one that often, suspiciously enough, turns out to be congruent with a nation-state. As a mediating force, “the social” tends to act as a fixed unity, like a lens; as a vocal force, as we’ve seen, it can do little more than to voice a putatively unanimous collectivity.

This is part of the reason so many ethnomusicologists have tended to prefer to approach voice as relational, even explicitly laying claim to what is sometimes called *relational ontology*. In brief, these approaches prioritize relations rather than sovereign entities or substances (Benjamin 2015). Anglophone and Francophone theorists of relationality tend to draw inspiration from Bakhtin (1981), Levinas ([1961] 1979), and Barad (2007); the tension between relational and substantialist ontologies also has a long tradition in Sanskritic metaphysics, often articulated as a tension between self-dependent being (*svabhāva*) and interdependent co-arising (*pratītyasamutpāda*; Dattreya [10th c.] 2018, 16; Nāgārjuna [3rd c.] 1987, 81).

Under the umbrella of “relationality” live a wide range of radically distinct analytic orientations. Some figurations, for example, stress the liminality of voice, which is seen to occur at the “boundaries” of discursive formations (Feldman 2015, 658). Others foreground vocal ensemble, demonstrating that certain kinds of vocality emerge only from joined voices (never from an individual voice), as, for example, in hocketing, counterpoint, or Kofi Agawu’s general account of “an irreducible togetherness” in Ewe song (Agawu 1995, 32; 2016, 121). Historical and ethnographic accounts of voice cultivation often emphasize the inherent cultivatedness of voice, trained to be what it is by caretakers, teachers, and exemplars, so that it is “paramparic” rather than self-produced (Herbst 1997; Rahaim 2012). Approaches that emphasize communication figure voicing as inherently directed toward a listener, or even as intermaterial vibration between a vocalizer and a listener (Kreiman and Sidtis 2011, 7; Eidsheim 2015, 3). Analyses of vocal interaction emphasize the improvised temporal *attunement* among interlocutors (Schutz [1932] 1967; Stone 1981). A more ontologically explicit approach, which we will explore later, focuses on the situated disclosure of voices, by which voices are constituted by disciplined listeners (Rahaim 2019). Each of these approaches to “relationality” is distinct. For example, Agawu is able to claim without contradiction that the irreducibly communal ethos of African vocal ensemble practices also implies a diminished attunement

to listeners (Agawu 2016, 120). Likewise, the sense in which a voice signifies a social totality or is found at discursive boundaries (as in Martha Feldman's figuration) begins from a sense of a voice with an inherent positionality, already stably located somewhere on a grid of signification, rather than unfolding in time or disclosed by situated practices of attunement.

Here, we will focus on just two of these (vocal interplay and the situated disclosure of voices), to return to our earlier theme of presence and temporality from another perspective.

THE TEMPORALITY OF VOCAL INTERPLAY

In Derrida's figuration of Husserlian voice, colloquy taints pure simultaneous self-presence with the inevitable, split-second absences of indication, retention, and interpretation. But in practice, colloquy relies far more richly on temporal play than just interpretive cognition. Vocal relations depend on rhythmic interaction on the scale of seconds. When I improvise a harmony over my friend's melody, I am anticipating where her voice is going based on what she has just sung; we are both mutually adapting (always imperfectly) to each other's rhythmic pulse, adjusting our volumes, finding our way into mutual tune, melding our vowels.

Though Derrida claims that Husserl's early account of vocal immediacy amounts to the "path down which . . . all of phenomenology has been pushed" (Derrida [1967] 2011, 3), this is certainly not obvious from the phenomenological literature. Most subsequent works in the tradition (e.g., Stein [1917] 1989; Schutz [1932] 1967; Merleau-Ponty [1945] 2012; de Beauvoir [1947] 1962; Levinas [1961] 1979) grapple with sociality, rapport, and contingent historical formations, rather than the absolute self-presence of a solitary, transcendental ego. Husserl himself seems to have intuited that an account of intersubjectivity was necessary to a well-rounded phenomenology ([1931] 1960, 89). The ambiguities that Derrida highlights may well be a result of the fact that Husserl's key work on temporality was gathered and edited from a series of orally delivered lectures in which he was working through problems that he never fully resolved. Husserl's reliance on a punctual melodic paradigm for his account of temporality may well have been a consequence of his rather tightly circumscribed musical education, which seems to have left him with a vision of music founded on finished, organically whole works, embodied in scores, rather than in performances (Ferrara and Behnke 1996, 468). Indeed, for an ethnomusicologist attuned to sociality, the most striking thing about the "music" of Husserl's temporality is that it is utterly asocial. It seems to simply *exist*, without any musicians, without purposive breath or gesture, without the disciplined practices of listening that disclose noemata in the first place, and without any sort of relationality. Jitendra Nath Mohanty, perhaps Husserl's foremost interpreter, considers this emphasis on soliloquy to be an incidental misstep on Husserl's part, pointing out

that phenomenology could not possibly proceed by reducing speech to private soliloquy; “on the contrary,” he writes, “it would seem to need ‘sharability’ and ‘communicability’” (Mohanty 1974, 243), as would, no less, an ethnomusicological phenomenology that foregrounds relationality.

The classic phenomenological analysis of relational temporality was offered by Alfred Schutz ([1932] 1967), whose account has been foundational for many ethnographers of performance (e.g., Geertz 1973; Stone 1981; Berger 2010; Friedson 1996; Feld 1988).¹³ In Schutz’s sociology, relationality is never simply a given fact like mass or electrical charge; it is accomplished in time. I may well share citizenship or convictions with millions of contemporaries that I’ve never met, but social intercourse with the *consociates* that I meet face to face requires something decidedly more musical (Schutz [1932] 1967, 8). Having a conversation, building a groove, taking a solo, singing in harmony, uttering affirmative *uh huhs* and *yeahs* and *mmms*, and the elaborate improvised choreography of conversational turn-taking—all of these are acts that require the ongoing, virtuosic negotiation of temporality with others: “mutual tuning-in” (Schutz 1976, 161). On this basis, Schutz posits a sociological distinction between the causal linkage of mutually anonymous contemporaries in a social system and the “mutually *interlocked*” actions that intersubjectively link consociates in a “community of time.” Schutz’s term for the latter (*aufeinander eingestellt*, “mutually adjusted” or “mutually attuned”) emphasizes the temporal dynamism of this kind of relationality (Schutz [1932] 1967, 180). For this constantly shifting attunement, mere coexistence is not enough; nor is semiotic expression or indication. Instead, interlocutors “follow [each] others’ action in its ongoing flux as it unrolls phase by phase.” Schutz specifies:

When I look at my consociate’s ongoing action, by protentions and anticipations I may expect its outcome even if I do not know his underlying project. . . . I participate in the ongoing flux of his action as directed toward its terminus, the goal to be attained, the act to be accomplished, the problem to be solved, the state of affairs to be brought about. (1996, 63)

This distinction between those with whom we interact and those absent multitudes with whom we merely coexist may seem to land us back in a Derridean metaphysics of presence. But the temporality that Schutz identifies in consocial interaction is radically different from the instantaneous self-presence of Husserlian-Derridean soliloquy. Social relations, unlike the Derridean cartoon of “presence,” could never happen in the oneness of solitude or the atemporal now-point marked by “the glance of an eye.” Doubly to the contrary, they unfold in a community of time. In such a community, retention and protention, rather than heralding absence, are precisely what make face-to-face sociality possible. Thus, attending to the temporality of voice is not an *alternative* to social analysis; temporality is a crucial *condition* of social relation.

Numerous ethnographic studies foreground the formation of contingent communities of time in mutually tuned-in vocal interaction. Kpelle group song, which features extensive hocketing and interlocked ostinato, yields a distinctive “inside” and “outside”

of a performance, and a sense of “going down the same road” when everything fits together properly (Stone 1981, 199–200). The interaction of Hindustani vocalists with their accompanists requires split-second reactions and fine mutual adjustment (Clayton 2007). The Kaluli practice of *dulugu ganalan*, in which voices overlap in exquisite cascades, brings singers into “non-hierarchical yet synchronous, layered, fluid group action” (Feld 1988, 83). Conversational interaction too relies on often-unconscious processes of rhythmic isochrony, whereby an elocutionary groove is reinforced by rhythmic interjections and turn-taking (Auer et al. 1999, 58–59). Each of these requires a finely disciplined temporal sense tuned for one particular form of ensemble.

Equally important are moments when ensembles break down, when voicely social actors “go down different roads” (Stone 1981, 200) or deliberately thwart established grooves. Kisliuk describes a particularly rich session of BaAka Pygmy song and dance performance where a separate men’s group and women’s group (each of which has members tuned in to each other but which, at the level of the two groups, are not coordinated) vie for control of which groove will come next (Kisliuk 2000, 30–37). In conversation, an interlocutor may deliberately break from a pulse, placing a stressed syllable slightly earlier or later than expected, in order to assert a contrary position (Auer et al. 1999, 80). As Emmanuel Levinas reminds us, rupture in conversation is at least as ethically productive as the “charm of rhythm” that pulls interlocutors toward participatory oneness ([1961] 1979, 203). These strategic temporal ruptures remind us that unity is just one part of social relations; alterity too is relational. But in all cases—in the cultivation of a participatory whole or in its contestation, in the formation of a temporal “bubble” or in its rupture—relationality depends on disciplines of retention and protention, rather than an instantaneous solipsistic immediacy.

VOCAL MULTIPLICITY

We have seen that attending to the temporality of vocal interaction may allow us to liberate the voice from the Husserlian-Derridean “metaphysical box.” If this box seemed convincing in the first place—if Derrida’s analysis of Husserlian soliloquy has seemed for the last fifty years to apply seamlessly to colloquy, song, political will, and public pronouncement—it surely is because “voice” is so often taken to be a coherent metaphysical category in the first place. The Voice (if there were such a universal, singular, transcendental entity) might appear to be paradoxical, everything and nothing at once: viscerally, materially *there*, yet invisible and intangible; at times distinctively individual (“she has finally found her voice”) and at others inescapably collective (“the people have spoken”); at times an outward expression of an active subjectivity resisting oppression (“speak truth to power!”), at others the mark of a passive, malleable subject disciplined by forces beyond its control (“learn to speak properly!”). This wild multiplicity is sometimes tamed under the sign of a single chimerical entity (The Voice) meant, in principle, to be all these things at once.

But the diverse entities that are placed under this category are as ontologically distinct as a handshake and a milkshake, despite the umbrella term “voice.” The “voice” of a baby crying, the “voice” of reason, the “voice” of a televangelist, the “voice” of the French people—these are not just acoustically different, but ontologically distinct, which is to say that they *are* differently. These are not merely different perspectives on a single, prior vocal reality but tokens of various kinds of realness. They have no common ground of being: not acoustics, not physiology, not lexical meaning. The phenomenological task—using a method I have elsewhere called *practical ontologies of voice*¹⁴—is to reflect on the situated processes by which voices show up in the first place. The result is not a unitary theory of The Voice but something like what, following Annemarie Mol’s brilliantly trod path (2002), we might call “The Voice Multiple.” We begin not from the abstract category of “The Voice” or from a putative lowest common denominator (like soliloquy, vibrations, or personhood) but from the concrete, practical, situated actions by which voices become so palpably real in the first place. This is the task of vocal ethnographies and vocal histories, like the one I am about to hazard now.

We begin in a voice clinic in Bombay—the clinic of Dr. Sadhana Nayak, a well-known voice therapist. I’m sitting with one of her patients, Apoorva.¹⁵ Apoorva is a versatile young singer of what in India is often called “light” music: romantic film songs, devotional bhajans, and the elaborate traditions of art song known as *ghazal* and *thumri*.¹⁶ Her singing voice is bright and strong, every tone encased in a delicate, numinous nasal resonance.

I sought out Apoorva not only for her evident skill in singing but also to hear her story—a story about losing and regaining her voice. She’s a compelling storyteller. Apoorva speaks in an even cadence; even when switching between Hindi and English midsentence, her syllables amble on without breaking stride. Once she gets into a groove, each sentence arcs downward in an attitude of a gentle resignation, without hedging, making its way to an inevitable tonic conclusion rather than soliciting approval or assent. Though she is speaking *to me*, answering my questions, there is no evident sense of feigning or fronting, of speaking for the sake of giving a certain impression. Following her from one point to another in her “talerealm” (Young 1987), I seem to simply hear *Apoorva* speaking, not *Apoorva’s voice* as a discrete third entity between us. Her story has its own flow, and I get into it.

So here’s her story. Four years previously, just as she had begun performing publicly, Apoorva was in a car accident that knocked her unconscious for a month and a half. When she woke up in the hospital, she had trouble coordinating the sides of her body; her right arm and leg in particular were sluggish and unresponsive. Walking, she said, became “harder than ballet.” Her speaking voice sounded alien: low and hoarse and breathy. Worse, when she tried to sing, there was, as she put it, “no voice” at all—only a whisper.

That’s when she went to see Dr. Nayak. Dr. Nayak numbed Apoorva’s throat and inserted a tiny camera and strobe light to render visible the motion of her vocal folds. She soon saw the malfunction she was looking for. Apoorva’s right vocal fold was hardly moving at all, apparently paralyzed from the neurological trauma. She turned off her

instruments, sat Apoorva down, and gave her exercises to do, as a physical therapist might: repetitive, daily vocal exercises far removed from the luminous tone and subtle poetics of Apoorva's usual vocal life.

The drudgery paid off. After only a week of daily practice, Apoorva found some improvement. She said that she could hold a tone—for just a moment, less than a second. But it was only after several more weeks of practice that “the voice started coming out.” Over the next few months, Apoorva's voice fully returned. It is now stronger than ever, and she's gone back to a successful career singing.

We now turn to reflection on the very different sorts of “voice” that we just encountered. First, consider the vocal capacity that Apoorva lost when she could speak but not sing. Here, the “voice” that is lost and later regained is not mere sound, nor mere vibrations. The whispery sound, though certainly vibratory and audible, is not it. Neither are the brief, tentative tones she made after a week (despite the fact that phonologists would call these utterances “voiced”). It is only the sustained, steady sound that came out after several weeks of practice that qualify as “*the voice coming out.*”

The “voice” here, lost for a time and later recovered through practice, is not mere sound, nor mere vibrations, nor even the use of the vocal cords. It is the capacity for a clear, stable, tuneful tone, or what is often called *sur*. This tone, like all vocal utterances, is not characterized by an atemporal, perfectly unchanging frequency. It is not simply given in the acoustical signal (in which there is always a measurable fluctuation of several hertz) but discerned by a listener with a trained ear who recognizes *sur* when they hear it. This requires an attunement to steady, sustained notes: we don't sit there counting oscillations (nor could we, without the technological prosthesis to disclose the waveform visually). A spoken sentence does not ordinarily show up as in tune or out of tune; attunement to singing is co-constitutive with a *singing* voice. Further, recall that when Apoorva woke up from her coma, she could speak but not sing. This is a problem reported by other traumatized singers; speaking poses no problem, and yet *sur* will not come. This is, I think, a suggestion of just how radically different *sur* is from mere speaking, akin to the difference between grasping and pointing described by Merleau-Ponty ([1945] 2012, 106).

A rather different “voice” is the vocal mechanism that was damaged in the crash. Dr. Nayak's office maps out this mechanism in boldly drawn anatomical diagrams of the lungs, tongue, and larynx, a dizzying labyrinth of tissues depicted in black and white, red and blue. A computer screen offers close-up footage of a particular pair of vocal folds, glistening with mucus. The laryngoscope, the strobe light, the anatomical illustrations hanging on her wall—these all disclose a world of interwoven muscle, cartilage, fluid, and bone that otherwise would be hidden away behind the skin of the throat. When Dr. Nayak works with a client to fix damage done to their “voice,” she is not referring to *sur*, or to a particular form of personhood, or to the voice that *is*, in some sense, Apoorva's story. The voice disclosed in the clinic is not sincere or insincere, true or false, in tune or out of tune. It is a complex organic mechanism made of flesh, which sustains damage, wears down, recovers, and—in this case, we hope—can be repaired.

Here, the guiding image of the laryngeal mechanism is the vivid, complex squirming of the living vocal folds. Seeing this is a rather different matter, practically, than counting the number of teeth in a patient's mouth. It's hard enough to see down there. But the squirming itself, unlike the teeth or tonsils, is not given to our senses at all. The slowest vocal vibrations—at dozens of cycles per second—are far too fast for the naked eye to resolve. Laryngeal stroboscopy uses a strobe light, in much the same way as a strobe-tuned phonograph reveals deviations from 33 1/3 RPM. A strobe light is tuned close to, but not precisely at, the frequency of the vibrations of the fold, offering a video image of the folds undulating like a stingray. The laryngoscope, the strobe light, the live feed on our computer screen, the careful tuning and focusing of each instrument until it looks the way we want it to—only with all of this in place is the living, moving mechanism of the voice disclosed visibly. (Still more sophisticated are the diagnostic techniques of laryngological seeing that disclose this squirming as symptomatic of health or disease.)

The third “voice” we encountered is that of Apoorva's story. Once we are with her in the flow of events, we are not faced with a sequence of words, nor a sequence of pitches. Indeed, as anyone who has transcribed an interview can attest, accessing these features after the fact requires real perceptual work. Nor even does her utterance show up as a set of historical *claims* to be challenged. Later, in reviewing the recording, I did indeed have to painstakingly reconstitute her utterances as a text, rewinding, slowing down parts where an individual word was lost, interpreting and transcribing pauses and inflections. I likewise constructed a timeline that I aligned with her doctor's account of the treatment. But in the moment of telling, the narrative texture of the talerealm is not marked for me as logically true as opposed to false. Nor is her sincerity a conclusion that I arrive at deductively from the words she says. Her story, her stance, and her person are bundled together in each utterance. I trustingly tune in. Don Ihde puts this beautifully:

When I listen to an other I hear him speaking. It is not a series of phonemes or morphemes which I hear, because to “hear” these I must break up his speech, I must listen “away” from what he is saying. My experiential listening stands in the near distance of language that is at one and the same time *the other speaking* in his voice. *I hear what he is saying.* ([1976] 2007, 151)

Apoorva's sincerity, in other words, is relational, unfolding in the temporality of our interaction. It has been constituted in part by my own practices of listening: trusting, attentive, patiently following the arc of her story on the scale of minutes, tuned into the temporality of her phrases on the scale of seconds. Nor is this a silent perceptual matter. From the moment we are introduced, I find myself encouraging her telling with my own *mhms* and *acchās*,¹⁷ offering not so much *logical* but *rhythmic* affirmation: mutual tuning in (see Figure 13.4).

By tuning in, by encouraging and affirming her telling, I become a participant in the story. While mutually engaged, it shows up to me as a more or less transparent report of what happened, not as a shrewd elicitation of sympathy or a set of claims to be challenged. The reliability of her sounding voice is not marked for me as logically true as

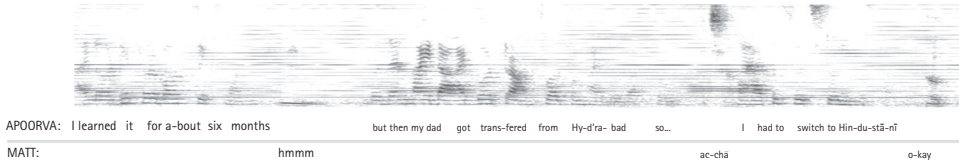


FIGURE 13.4: Mutual tuning in.

opposed to false; nor is it a conclusion that I arrive at step by step, through deliberative reason. Such an appearance of sincerity may seem, in contrast to the suspiciously artificial persona of a poseur or a salesman, to be “im-mediate,” that is, not mediated by an ulterior front between her and me. And yet sincerity itself is not an inherent feature of her voice to which I have gained direct access. Nor, even, is its status as a “voice” given in the acoustical signal. It is achieved in our mutual vocal performance. This improvised, provisional, unfinished intersubjective present that we share is by no means an unmediated metaphysical presence. We are mutually tuned in to the arc of each other’s sentences, retaining what came before, protending their conclusions, timing our turn-taking in accordance with this empathy.

Each of these three “voices”—sur, mechanism, story—is disclosed through a particular stance, through a particular technique of listening, through a particular infrastructure. Tuning into sur, physiology fades into the background; getting into the flow of a story, sur fades into the background; attending to the details of physiology, narrative fades into the background. The referential span of these *voice*-s may well be a source of wonder, of curiosity, of humility. But The Voice as an abstract theoretical category has no *svabhāva*, no self-dependent being. Any of these phenomena called voice emerge only in *pratīyasamutpāda*: in a web of ongoing, interdependent interbeing.

VOCAL INTIMACIES AND IMMEDIA

All of this is not to say that laryngeal function, or in-tuneness, or sincerity, or any other vocal feature is not *real*. On the contrary, it is to affirm multiple (if not necessarily infinite) vocal realnesses. These are not mere personal subjective flights of fancy, whereby one hallucinates voices at will. That they are not arbitrary is evidenced by the very fact that each of these ontological formations affords surprise and rewards close investigation with more detail. (Indeed, even auditory vocal hallucinations, which are often dismissed as “all in the head,” typically show up as being out there in the world.)¹⁸ This is why it is possible for a group of properly enculturated and attuned vocalists to discern and correct an out-of-tune voice in a harmonic texture without any arithmetical calculations, for two laryngologists to collaboratively diagnose a vocal malfunction, or for a panel of judges to agree on which recitation of a sura from the Qur’an is the best. Even vocal impersonators succeed or fail on the evidence of how *real* their vocal

“falsehood” is. Each of these ontological formations of voice has its own durable reality, internally consistent, even if mutually incommensurable with the others. Practical ontology foregrounds the skillful techniques that disclose each durable realness in its turn—or, at times, indeterminately, as two or more at once (Rahaim 2019, 30–31).

Various vocal immediacies are implied by these realnesses. For example, a live radio broadcast is (in a limited sense) “immediate” in comparison to one that is prerecorded; speech in person is “immediate” in comparison to a live radio broadcast; a spontaneous utterance is “immediate” in comparison to a memorized one; a song in its original language is “immediate” in comparison to a translation. None of these immedia, however, imply the immaculate metaphysical self-presence that Derrida diagnoses in Husserlian soliloquy. The inevitable “mediations” and “absences” of these contingent immedia— anonymity, temporal delays, durable habits, great distances—do not amount to a single metaphysical substance. Indeed, phenomenological reflection reveals that the “mediation” of, say, phonetic distinction has nothing essentially in common with the “mediation” of linguistic translation or the “mediation” of electronic transduction. Nor, indeed, does “immediacy” imply a lack of media (Kim 2016, 10). The palpable intimacy of Kishore Kumar’s close-miked crooning, of Björk’s audible intake of breath between phrases, of Zeki Müren’s velvety tone, which reaches out to listeners in the privacy of a living room or a car, or even the “privatized auditory bubble” of headphones—all of these are made possible through sophisticated techniques of microphonic transduction, amplification, and reproduction (Stokes 2010, 61; Bull 2005, 344). Likewise, the palpable “immediacy” of face-to-face conversation or improvised polyphony could never happen *immediately*. Even the most intimate vocal presence unfolds in time, is disclosed through disciplined practices of listening and voicing, depends on spaces and infrastructures and technologies—and even what we conventionally call “media.” Thus, simply invoking absence is not enough. The key phenomenological task for ethnographers is to account for the situated and widely varied conditions and practices that disclose a voice in the first place. This task is especially urgent when a voice seems to simply be there, self-sufficient, self-identical, self-expressive, in the blink of an eye.

NOTES

1. The book’s original title is *La voix et le phénomène*. The previous translation of the title as *Speech and Phenomenon* likely was to clarify the rather limited scope of “voice” at issue in the book; yet Derrida’s extended critique of logocentrism nonetheless continued to thematize voice as such, as, for example, in his treatment of the neume in *Of Grammatology* ([1967] 1997, 249).
2. See, for example, Charles Taylor (1989, 362) and Mladen Dolar (2006, 87) on “inner voice” as a voice of truth, in contrast to the compromises and falsehoods of public voices.
3. In Hume’s case ([1738] 1960, 36), the notes are tooted out on a flute.
4. An actual pianist playing a piano, of course, also is involved in gestural melody; their hands are always in motion, tracing arcs in space.

5. Hindustani vocalists refer to scale degrees in *sargam* syllables, akin to movable do solfege—*sā re ga ma pa dha ni*. In curly brackets, I have translated *sargam* into scale-degree numbers.
6. Most singers of raga music have a mastery of *svar-jñān* (the capacity to discern note names by ear), to be sure, but not all singers of kirtan, qawwali, film music, or pop music do. See Neuman (2004) and Rahaim (2012) for an account of raga singers who can sing fluidly without notation.
7. Elsewhere, in a discussion of the intuition of “incomplete” melodies, Husserl does seem to invite a conception of melody as motion ([1928] 1991, 141–155).
8. Sakhawat Hussain Khan’s original Hindi phrase is *gāyan kaimara* ([1952] 1976). Max Katz’s sensitive translation of this phrase (2012) correctly renders it as “singing camera,” which for our present, nitpicking philosophical purposes may read ambiguously. To be precise, *gāyan* operates here as an adjectival noun, as in *gāyan samāj* (a connoisseur’s society to support the art of singing, not a society that sings); notation is thus construed as a *camera for singing* (fixing song into punctual, frozen moments), not a *camera that sings*.
9. Zahavi explicitly points out that the primal impression is “Husserl’s name for the consciousness of the now-phase of the [temporal] object, and not the name for this now-phase itself” (1999, 65).
10. Hubert Dreyfus (1995, x) points out that the term *Augenblick* has a specifically Lutheran heritage, translating the biblical “twinkling of an eye” in which “we shall be changed” (1 Corinthians 15:52).
11. To be precise, Derrida translates *Augenblick* as *clin d’oeil*, which ambiguously indicates both blinking and twinkling. He then exploits the pun by explicitly thematizing the closed eyelid of blinking, which is not implied by Husserl’s original *Augenblick*.
12. See Latour (2005, 8–14) for a discussion of this distinction in sociology. There, he calls the collectivist approach the “sociology of the social” and the relational approach the “sociology of associations.”
13. It’s worth noting that in his application of Schutz’s ideas, Clifford Geertz misreads the notion of a “community of time” to mean merely contemporaries on the scale of months or years and thus misses the rhythmic specificity of consocial interplay (1973, 364–365).
14. This term was coined by Gad et al. (2015; see also Rahaim 2019). Another common use of the term “ontology” simply refers to claims about what *is*. For example, we might say that a “materialist ontology” posits benches and bricks, but not fairies; we might say that a “relational ontology” sees a world in which relations are prior to discrete entities. I distinguish between these attempts to enumerate what exists from a *practical ontological* account, which reflects on the being of these entities, and *how* these worlds of entities are disclosed.
15. In this chapter, I have left out her last name to maintain her anonymity.
16. Though these genres are often characterized in the Hindustani tradition as types of “semi-classical” music (as compared with supposedly “classical” raga music forms such as *khyal* and *dhrupad*), ghazal and thumri are classical vocal forms in all of the usual senses of that term. Both ghazal and thumri draw on vast poetic, modal, and melodic canons; both have centuries of literate as well as oral tradition; and both are rooted in courtly milieu, aristocratic patronage, and aesthetic criticism.
17. *Acchā* is a Hindi expression meaning “good” or “I get it.”
18. Describing the experiences of a patient with auditory hallucinations, psychologist Eugène Minkowski observed, “In the street, it is like a whispering that *envelops him entirely*; in the same way he feels deprived of his freedom as if *around him* people were always present; at the café, it is as if there is something nebulous *around him* and he senses a trembling; and

when the voices are particularly frequent and numerous, the atmosphere *around him* is saturated as if by fire” (Minkowski [1932] quoted in Merleau-Ponty [1945] 2012, 545n77); italics reproduced from Merleau-Ponty).

WORKS CITED

- Agawu, Kofi. 1995. *African Rhythm: A Northern Ewe Perspective*. Cambridge: Cambridge University Press.
- Agawu, Kofi. 2016. *The African Imagination in Music*. New York: Oxford University Press.
- Aldererson-Day, Ben, and Charles Fernyhough. 2015. “Inner Speech: Development, Cognitive Functions, Phenomenology, and Neurobiology.” *Psychological Bulletin* 141 (5): 931–965.
- Auer, Peter, Elizabeth Couper-Kuhlen, and Frank Muller. 1999. *Language in Time: The Rhythm and Tempo of Spoken Interaction*. New York: Oxford University Press.
- Bakhtin, M. M. 1981. *The Dialogic Imagination: Four Essays*. Edited by Michael Holquist. Translated by Caryl Emerson and Michael Holquist. Austin: University of Texas Press.
- Barad, Karen. 2007. *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durham, NC: Duke University Press.
- Beauvoir, Simone de. (1947) 1962. *The Ethics of Ambiguity*. Translated by Bernard Frechtman. Secaucus, NJ: Citadel Press.
- Benjamin, Andrew. 2015. *Towards a Relational Ontology: Philosophy’s Other Possibility*. Albany: State University of New York Press.
- Berger, Harris. 2010. *Stance: Ideas about Emotion, Style, and Meaning for the Study of Expressive Culture*. Middletown, CT: Wesleyan University Press.
- Bergson, Henri. [1889] 2001. *Time and Free Will: An Essay on the Immediate Data of Consciousness*. London: Sonnenschein.
- Brentano, Franz. [1874] 1995. *Psychology from an Empirical Standpoint*. London: Routledge.
- Brubaker, Rogers. 2002. “Ethnicity without Groups.” *European Journal of Sociology* 43 (2): 163–189.
- Bull, Michael. 2005. “No Dead Air! The iPod and the Culture of Mobile Listening.” *Leisure Studies* 24 (4): 343–355.
- Cavarero, Adriana. 2005. *For More than One Voice: Toward a Philosophy of Vocal Expression*. Stanford, CA: Stanford University Press.
- Clayton, Martin. 2007. “Time, Gesture and Attention in a ‘Khyāl’ Performance.” *Asian Music* 38 (2): 71–96.
- Danielson, Virginia. 1998. *“The Voice of Egypt”: Umm Kulthūm, Arabic Song, and Egyptian Society in the Twentieth Century*. Chicago: University of Chicago Press.
- Dattreya. (10th c.) 2018. *Avadhuta Gita: The Song of the Avadhuta*. Translated by Janki Parikh. Self-published. <http://www.jankiparikh.com/>.
- Derrida, Jacques. (1967) 1997. *Of Grammatology*. Translated by Gayatri C. Spivak. Baltimore, MD: Johns Hopkins University Press.
- Derrida, Jacques. (1967) 2011. *Voice and Phenomenon: Introduction to the Problem of the Sign in Husserl’s Phenomenology*. Translated by Leonard Lawler. Evanston, IL: Northwestern University Press.
- Desai-Stevens, Anaar. 2017. “Singing through the Screen: Indian Idol and the Cultural Politics of Aspiration in Post-liberalization India.” PhD diss., Cornell University.

- Deva, B. C. 1959. "The Vibrato in Indian Music: Studies in Indian Musical Scales II; Gamakas." *Acustica* 9: 175–180.
- Dolar, Mladen. 2006. *A Voice and Nothing More*. Cambridge, MA: MIT Press.
- Dreyfus, Hubert. 1995. *Being-in-the-World: A Commentary on Heidegger's "Being and Time, Division I"*. Cambridge, MA: MIT Press.
- Eidsheim, Nina. 2015. *Sensing Sound: Singing and Listening as Vibrational Practice*. Durham, NC: Duke University Press.
- Faubion, James. 2011. *An Anthropology of Ethics*. Cambridge: Cambridge University Press.
- Feld, Steven. 1984. "Sound Structure as Social Structure." *Ethnomusicology* 28 (3): 383–409.
- Feld, Steven. 1988. "Aesthetics as Iconicity of Style or 'Lift-up-over Sounding': Getting into the Kaluli Groove." *Yearbook for Traditional Music* 20: 74–113.
- Feldman, Martha. 2015. "Why Voice Now?" *Journal of the American Musicological Society* 68 (3): 653–685.
- Ferrara, Lawrence, and Elizabeth Behnke. 1996. "Music." In *Encyclopedia of Phenomenology*, edited by Lester Embree et al., 467–473. Boston: Springer.
- Friedson, Steven M. 1996. *Dancing Prophets: Musical Experience in Tumbuka Healing*. Chicago: University of Chicago Press.
- Frith, Simon. 1998. *Performing Rites: On the Value of Popular Music*. Cambridge, MA: Harvard University Press.
- Gad, Christopher, Casper Bruun Jensen, and Brit Ross Winthereik. 2015. "Practical Ontology: Worlds in STS and Anthropology." *NatureCulture*, no. 3, 67–86. <https://www.natcult.net/wp-content/uploads/2018/12/PDF-natureculture-03-04-practical-ontology.pdf>.
- Geertz, Clifford. 1973. *The Interpretation of Cultures*. New York: Basic Books.
- Harkness, Nicholas. 2013. *Songs of Seoul: An Ethnography of Voice and Voicing in Christian South Korea*. Berkeley: University of California Press.
- Herbst, Edward. 1997. *Voices in Bali: Energies and Perceptions in Vocal Music and Dance Theater*. Middletown, CT: Wesleyan University Press.
- Hume, David. (1738) 1960. *A Treatise of Human Nature*. Oxford: Oxford University Press.
- Husserl, Edmund. (1900) 2001. *Logical Investigations*. Vol. 1. London: Routledge.
- Husserl, Edmund. (1913) 1982. *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy*. Translated by F. Kersten. The Hague: Martinus Nijhoff.
- Husserl, Edmund. (1928) 1991. *On the Phenomenology of the Consciousness of Internal Time*. Translated by John Barnett Brough. Dordrecht, Netherlands: Kluwer.
- Husserl, Edmund. (1931) 1960. *Cartesian Meditations: An Introduction to Phenomenology*. Translated by Dorion Cairns. The Hague: Martinus Nijhoff.
- Ihde, Don. (1976) 2007. *Listening and Voice: Phenomenologies of Sound*. 2nd ed. Albany: State University of New York Press.
- Kane, Brian. 2015. "The Model Voice." In "Colloquy: Why Voice Now?" *Journal of the American Musicological Society* 68 (3): 671–676.
- Katz, Max. 2012. "Institutional Communalism in North Indian Classical Music." *Ethnomusicology* 56 (2): 279–298.
- Kendon, Adam. 2004. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
- Khan, Sakhawat Hussain. (1952) 1976. "Gayan Camera: Notation." In *Swarna Jayanti Smarika* [Golden jubilee souvenir], edited by Puru Dadheech, 34–35. Lucknow: Bhatkhande Hindustani Sangeet Mahavidyalaya.

- Kim, John. 2016. *Rupture of the Virtual*. DigitalCommons@Macalester College. <https://digitalcommons.macalester.edu/books/1/>.
- Kisliuk, Michelle. 2000. "Performance and Modernity among BaAka Pygmies: A Closer Look at the Mystique of Egalitarian Foragers in the Rain Forest." In *Music and Gender*, edited by Pirkko Moisala and Beverley Diamond, 25–50. Urbana: University of Illinois Press.
- Kreiman, Jody, and Diana Sidtis. 2011. *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception*. Chichester, UK: Wiley-Blackwell.
- Kunreuther, Laura. 2014. *Voicing Subjects: Public Intimacy and Mediation in Kathmandu*. Berkeley: University of California Press.
- Latour, Bruno. 2005. *Reassembling the Social: An Introduction to Actor-Network Theory*. New York: Oxford University Press.
- Levinas, Emmanuel. (1961) 1979. *Totality and Infinity: An Essay in Exteriority*. Translated by Alfonso Lingis. Pittsburgh, PA: Duquesne University Press.
- Lomax, Alan. 2003. *Selected Writings, 1934–1997*. Edited by Ronald D. Cohen. New York: Routledge.
- Lorde, Audre. 2020. *Sister Outsider: Essays and Speeches*. New York: Penguin.
- Luhrmann, Tanya. 2012. *When God Talks Back: Understanding the American Evangelical Relationship with God*. New York: Alfred A. Knopf.
- McNeill, David. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.
- Merleau-Ponty, Maurice. (1945) 2012. *The Phenomenology of Perception*. Translated by Donald A. Landes. New York: Routledge.
- Minkowski, Eugène. 1932. "Le problème des hallucinations et le problème de l'espace." *Évolution psychiatrique* 2 (3): 57–76.
- Mohanty, Jitendra Nath. 1974. "On Husserl's Theory of Meaning." *Southwestern Journal of Philosophy* 5 (3): 229–244.
- Mol, Annemarie. 2002. *The Body Multiple: Ontology in Medical Practice*. Durham, NC: Duke University Press.
- Nāgārjuna. (3rd c.) 1987. *Nagarjuna's Seventy Stanzas: A Buddhist Psychology of Emptiness*. Translated by Tenzin Dorjee and David Ross Komito. Ithaca, NY: Snow Lion Publications.
- Nancy, Jean-Luc. (2002) 2007. *Listening: A History of Our Ears*. Translated by Charlotte Mandell. New York: Fordham University Press.
- Neuman, Dard. 2004. "A House of Music: The Hindustani Musician and the Crafting of Traditions." PhD thesis, Columbia University.
- Rahaim, Matthew. 2012. *Musicking Bodies: Gesture and Voice in Hindustani Music*. Middletown, CT: Wesleyan University Press.
- Rahaim, Matthew. 2019. "Object, Person, Machine, or What: Practical Ontologies of Voice." In *The Oxford Handbook of Voice Studies*, edited by Nina Eidsheim and Katherine Meizel, 19–34. New York: Oxford University Press.
- Schultz, Anna. 2012. *Singing a Hindu Nation: Marathi Devotional Performance and Nationalism*. Chicago: University of Chicago Press.
- Schutz, Alfred. (1932) 1967. *The Phenomenology of the Social World*. Translated by George Walsh and Frederick Lehnert. Evanston, IL: Northwestern University Press.
- Schutz, Alfred. 1976. *Collected Papers*. Vol. 2. The Hague: Martinus Nijhoff.
- Schutz, Alfred. 1996. *Alfred Schutz: Collected Papers*. Vol. 4. Dordrecht, Netherlands: Springer Science & Business Media.

- Srivastava, Sanjay. 2004. "Voice, Gender, and Space in the Time of Five-Year Plans: The Idea of Lata Mangeshkar." *Economic and Political Weekly* 39 (20): 2019–2028.
- Stein, Edith. (1917) 1989. *On the Problem of Empathy*. 3rd rev. ed. Vol. 3 of *Collected Works of Edith Stein*. Washington, DC: ICS Publications.
- Stokes, Martin. 2010. *The Republic of Love: Cultural Intimacy in Turkish Popular Music*. Chicago: University of Chicago Press.
- Stone, Ruth. 1981. "Toward a Kpelle Conceptualization of Music Performance." *Journal of American Folklore* 94 (372): 188–206.
- Streeck, Jürgen. 2009. *Gesturecraft: The Manu-facture of Meaning*. Philadelphia, PA: John Benjamins.
- Sudnow, David. 1978. *Ways of the Hand*. Cambridge, MA: MIT Press.
- Taylor, Charles. 1989. *Sources of the Self: The Making of the Modern Identity*. Cambridge, MA: Harvard University Press.
- Vygotsky, L. S. (1934) 1987. *Thinking and Speech: The Collected Works of Lev Vygotsky*. Vol. 1. New York: Plenum Press.
- Weidman, Amanda. 2011. "Anthropology and the Voice." *Anthropology News* 52 (1): 13.
- Weidman, Amanda. 2012. "The Ethnographer as Apprentice: Embodying Sociomusical Knowledge in South India." *Anthropology and Humanism* 37 (2): 214–235.
- Young, Katharine Galloway. 1987. *Taleworlds and Storyrealms*. Dordrecht, Netherlands: Martinus Nijhoff.
- Zahavi, Dan. 1999. *Self-Awareness and Alterity: A Phenomenological Investigation*. Evanston, IL: Northwestern University Press.