ALEX BORKOWSKI

Vocal Aesthetics, AI Imaginaries

Reconfiguring Smart Interfaces

ABSTRACT In 2019, a report commissioned by the United Nations Educational, Scientific and Cultural Organization (UNESCO) claimed that the way users interface with technology is in the midst of a paradigm shift—from text input and output to voice input and output. Indeed, voice assistants such as Amazon's Alexa, Apple's Siri, and Google Assistant are increasingly ever-present in domestic spaces. Against this backdrop, this paper proposes that critical thinking about Artificial Intelligence (AI) might be advanced and elaborated by simultaneously thinking critically about voice, unpacking beliefs that spoken language operates as a seamless "natural user interface," as well as broader ideological convictions about voice's exceptional status as a marker of unmediated self-presence. I examine the entangled conceptual roots between theories of voice and apocryphal origin stories of artificial intelligence in two eighteenth-century inventions—Wolfgang von Kempelen's Mechanical Turk and speaking machine. The aesthetic practices surrounding these and other so-called thinking and speaking machines engender shifting relations of visibility and invisibility, obfuscating the role of human labor in creating the seeming magic of AI. I also examine two artistic projects—Holly Herndon's electropop album *PROTO* (2019) and a series of collaborative works by Sofian Audry and Erin Gee titled *of the soone* (2018)*, to the sooe* (2018), and *Machine Unlearning* (2018–19)—that, in their varied deployments of machine learning in vocal performance, challenge the ideological foundations of the sociotechnical milieu from which they emerge. These artworks tune into the frictions and glitches inherent to vocal interfaces and thereby unravel the vocal imaginary that sustains the illusion of their neutrality, self-sufficiency, or, indeed, intelligence.

KEYWORDS artificial intelligence, digital assistance, voices, interfaces, media art, Holly Herndon, Sofian Audry, Erin Gee, Wolfgang von Kempelen

INTRODUCTION

Smart technology seems to be gripped by something of a fervor for voice. In 2019, a report commissioned by the United Nations Educational, Scientific and Cultural Organization (UNESCO) claimed that the way users interface with technology is in the midst of a paradigm shift—from text input and output to voice input and output.[1] Indeed, voice assistants such as Amazon's Alexa, Apple's Siri, and Google's Assistant—intelligent agents who respond to a variety of user queries and execute tasks within the broader network of the smart home—have only become increasingly central to daily life

1. Mark West, Rebecca Kraut, and Han Ei Chew, "I'd blush if I could: closing gender divides in digital skills through education," UNESCO and EQUALS Skills Coalition, United Nations Educational, Scientific, and Cultural Organization website, 2019, 95, https://unesdoc.unesco.org/ark:/48223/pf0000367416.

Afterimage, Vol. 50, Number 2, pps. 129–149. ISSN 2578-8531. DOI: https://doi.org/10.1525/aft.2023.50.2.129

in the intervening years as one of the most ubiquitous applications of natural language processing and synthesis.[2] "Voice is everywhere," proclaimed a report published in June 2022 by NPR and Edison Research; 62 percent of American adults interact with a voice assistant via their smart phones, in-car systems, or smart speakers, while 35 percent, around 100 million people, have at least one smart speaker in their home.[3] Researchers speculate that a "conversational internet, mediated not by a web browser but by a machine that listens and talks like a person" is poised to replace familiar visual and haptic interfaces.[4] Voicebot.ai, an effusive tech blog whose stated mission is "giving voice to a revolution," prophesizes that Alexa and Google Assistant will come to be understood as "tectonic shifts in computing," akin to the Apple iPhone's touch screen.[5]

While voice assistants are inarguably popular and increasingly ubiquitous, there is something curious about the rhetorical invocation of paradigm shifts and sensory revolutions—the aural poised to overcome the visual and haptic. Vocal interfaces have long been a part of the cultural imaginary, as evinced by the myriad conversational computers that appear in science fiction. Yet today's voice assistants are posited, by virtue of their vocal nature as such, as entirely new and disruptive. Amazon promises users that "by letting your voice control your world," Alexa unlocks "possibilities you've never imagined."[6] The Alexa Fund, a venture capital initiative supporting innovation in Artificial Intelligence (AI) and voice, is purportedly premised on a belief that "experiences designed around the human voice will fundamentally improve the way people use technology."[7] This framing of the turn to voice as novel and groundbreaking, a perspective advanced by tech companies, commentators, and researchers alike, might be understood alongside the tendency in scholarly perspectives to situate sound as a radical modality that offers a challenge to dominant visual paradigms. Jonathan Sterne has observed that the ways in which sound has been ascribed certain "natural" or transhistorical traits are in fact ideologically entrenched in Christian ontotheology.[8] Further, such perspectives advance narratives in which history emerges as a series of zero-sum sensory paradigm shifts; yet, Sterne argues, the claim that cultural change correlates to one sense becoming dominant over another has little empirical basis.[9] In keeping with this critique, as well as other recent scholarship that dispenses with the conviction that attending to sound is

2. The scope of this paper is restricted to audible voice; however, complementary discussions regarding authorial voice and artificial intelligence, as in text-based chatbots, are taking place elsewhere. See Avery Slater, "Chatbots: Cybernetic Psychology and the Future of Conversation," *JCMS: Journal of Cinema and Media Studies*61, no 4 (Summer 2022): 181–7.

3. "The Smart Audio Report," National Public Radio and Edison Research (Spring 2022), www.nationalpublicmedia.com/uploads/2022/06/The-Smart-Audio-Report-Spring-2022.pdf.

4. West, Kraut, and Chew, "I'd blush if I could," 96.

5. "Mission," Voicebot.ai, https://voicebot.ai/about.

6. "Alexa Features," Amazon, www.amazon.ca/gp/browse.html?language=en_CA&node=21497226011&tag=googcana20&hvadid=240759136309&hvpos=&hvexid=&hvnetw=g&hvrand=13211596429057213416&hvpone=&hvptwo=&hvqmt=e&hvdev=c&ref=pd_sl_752zwdtjtv_e&hvtargid=kwd-295921609170.

7. "The Alexa Fund," Amazon, https://developer.amazon.com/en-US/alexa/alexa-startups/alexa-fund.

8. Jonathan Sterne, *The Audible Past: Cultural Origins of Sound Reproduction*(Durham, NC: Duke University Press, 2003), 14–15.

9. Sterne, *The Audible Past*, 16.

necessarily revolutionary or recuperative,[10] this paper approaches adulatory claims regarding smart technology and voice incredulously. Rather than situating the contemporary technoscape as one in transition from text-based to voice-based interaction, I suggest that the very promise of such a transition itself constitutes a vocal imaginary, one laden with ideological baggage regarding communication, agency, and the parameters of human subjectivity.

As daily experience is ever more permeated with interfaces, voice provides a mode to insert yet another point of contact with digital platforms without displacing existing ones. Users can engage with Amazon and Google aurally, while simultaneously typing and scrolling; thus, this hands-free interaction afforded by smart speakers is posited as a way to eliminate friction in,[11] and introduce new channels for, consumers' engagement with the digital marketplace. In this respect, as one among many interfaces, voice assistants might be understood in accordance with Alexander R. Galloway's framing of the interface as a threshold device, a passage or point of transition, whose success is measured by its transparency: "for every moment of virtuosic immersion and connectivity . . . of inopacity, the threshold becomes one notch more invisible, one notch more inoperable."[12] Indeed, this logic is evident in claims that voice offers a more organic, transparent, and accessible mode of engaging with digital technology, as voice-operated devices are posited as qualitatively distinct from their visual counterparts—even *more* inoperable, achieving more by doing less, by virtue of their recourse to spoken language as an inherently "natural user interface."[13] Embedded in such propositions are assumptions regarding the innate or universal traits of voice, many of which are predicated upon a belief in voice's inherent "naturalness" or proper relationship to a human subject.[14] Potent metaphors abound regarding finding and giving voice, such that it is granted exceptional status as a marker of agency, self-possession, and unmediated self-presence. Voice is necessarily personal—stable, essential, and singular—as well as necessary to one's civic personhood; it is "the ticket to entrance into the human community."[15] Indeed, Amazon's rhetorical invocation of "the human voice" is striking, as one never hears technologies that involve manual typing or swiping described as experiences around "the human touch."

Voice assistants might therefore be understood as a sociotechnical nexus in which this vocal imaginary is entwined with what Claudia Schmuckli calls an "AI imaginary"—"the trove of images and symbols derived from the metaphors that guide and describe the

10. Robin James, *The Sonic Episteme: Acoustic Resonance, Neoliberalism, and Biopolitics* (Durham, NC: Duke University Press, 2019), 4.

11. Emily West, *Buy Now: How Amazon Branded Convenience and Normalized Monopoly* (Cambridge, MA: MIT Press, 2022), 134.

12. Alexander R. Galloway, *The Interface Effect* (Cambridge, MA: Polity Press, 2012), 25.

13. Thao Phan, "The Materiality of the Digital and the Gendered Voice of Siri," *Transformations* 29 (2017): 28.

14. For a more robust discussion of prominent ideological assumptions regarding voice, and their audist and ableist underpinnings, see Jonathan Sterne, *Diminished Faculties: A Political Phenomenology of Impairment* (Durham, NC: Duke University Press, 2022).

15. Dominic Pettman, *Sonic Intimacy: Voice, Species, Technics (or, How to Listen to the World)* (Redwood City, CA: Stanford University Press, 2017), 4.

design, operations, and applications of AI."[16] Such metaphors flourish precisely because of the slipperiness of the term "AI" itself[17]; it proliferates through cultural conversations regarding the power and possibilities of artificial intelligence—both utopian and dystopian—that rarely ask what precisely it is.[18] Meredith Broussard writes regarding popular representations and perceptions of AI that "it's easy to confuse what we imagine and what is real."[19] While users would certainly be mistaken in perceiving voice assistants' conversational abilities as genuine understanding, as Broussard notes,[20] such imaginings are "real" in the sense that they are operative in structuring day-to-day engagements with such technologies.[21] I therefore suggest that it is precisely the looseness of the term "AI" in its popular usage that facilitates its ideological power. It is the overall absence of a simple definition of AI that makes it possible for convictions about the innateness, intimacy, and authenticity of voice to become so readily wedded to smart technologies as a testament to their purported impartiality, efficiency, and accuracy. Given the copious evidence attesting to the coalescence of machine learning with surveillance capitalism, the biases embedded in algorithmic processes, and the ways that such seemingly immaterial tech is built upon natural resource extraction and exploited human labor, this paper therefore proposes that critical thinking about AI might be advanced and elaborated by simultaneously thinking critically about voice—how are these imaginaries mobilized in harmony with one another and to what ends?

This paper thus interrogates a synchronicity between theories of voice and an apocryphal origin story of AI, locating their entangled conceptual roots in the late eighteenth century. I examine the connections between Amazon's crowdsourcing platform MTurk and its namesake, a chess-playing automaton invented and styled by Wolfgang von Kempelen as the Mechanical Turk. Despite von Kempelen's claims that his creation was capable not only of autonomous movement but independent thought, the Mechanical Turk was in fact entirely controlled by a human operator. Relatedly, MTurk is a platform that distributes micro-tasks to a vast decentralized and underpaid workforce, such that seemingly automated systems are sustained by human labor. While this lineage is well documented, the pairing of the Mechanical Turk with von Kempelen's "speaking machine"—an earnest attempt to generate mechanical speech that was often presented

16. Claudia Schmuckli, "Automatic Writing and Statistical Montage," in *Beyond the Uncanny Valley: Being Human in the Age of AI*, exh. cat. (San Francisco: Fine Arts Museums of San Francisco, 2020), 9.

17. Sarah T. Roberts, "Your AI is A Human," in *Your Computer is on Fire*, ed. Thomas S. Mullany, Benjamin Peters, Mar Hicks, and Kavita Philip (Cambridge, MA: MIT Press, 2021), 52.

18. Yarden Katz, *Artificial Whiteness: Politics and Ideology in Artificial Intelligence* (New York: Columbia University Press, 2020), 3.

19. Meredith Broussard, *Artificial Unintelligence: How Computers Misunderstand the World* (Cambridge, MA: MIT Press, 2018), 31.

20. Broussard, *Artificial Unintelligence*, 38.

21. Indeed, voice assistants are posited in their design and branding as intelligent entities. Amazon selected the name Alexa as a reference to the Library of Alexandria, alluding to the depth of knowledge cultivated by the agent. See Alexa Juliana Ard, "Amazon, can we have our name back?," *Washington Post*, December 3, 2021, www.washingtonpost.com/technology/interactive/2021/people-named-alexa-name-change-amazon/?itid=lk_inline_manual_21.

as a credibility-lending prelude to the Mechanical Turk[22]—merits further examination. I therefore take up this historical antecedent in order to interrogate how voice contributes to the shifting relations of visibility and invisibility that continue to inform the aesthetic practices surrounding so-called "thinking machines."

While the pairing of the Mechanical Turk with the speaking machine is evidence of the ways that a vocal imaginary might be mobilized to lend credence to a purportedly intelligent machine, it simultaneously begins to unravel the humanist paradigms upon which such an imaginary relies. Indeed, this "double device" acts as a catalyzing anecdote for Mladen Dolar's theory of voice, which identifies a commonality between human and mechanical vocalizations as uncanny "effect[s] without a proper cause."[23] Attending to voice in this way—in its affective and extra-communicative dimensions—opens up a counter-discourse that disrupts the purported seamlessness afforded by vocal interfaces. This obverse framework, which emerges from and remains embedded in synthesized voices, establishes the theoretical ground upon which I will examine two projects in contemporary art and music—Holly Herndon's electropop album *PROTO* (2019) and a series of collaborative works by Sofian Audry and Erin Gee titled *of the soone* (2018), *to the sooe* (2018), and *Machine Unlearning* (2018–19). In their varied deployments of machine learning in vocal performance, these artworks challenge the ideological foundations of the sociotechnical milieu from which they emerge. While neither deals explicitly with the politics of voice assistants or intervenes directly in the hardware or software of smart speakers—as in the work of Wesley Goatley, Lauren Lee McCarthy, or Martine Syms—both work to subvert claims that voice comprises the most natural, and therefore most invisible, interface. These artworks linger in and with what Galloway identifies as the "intraface"—a "zone of indecision" between "the workable and unworkable" internal to the interface[24]—thereby deploying voice as a mode of digging into, rather than glossing over, the complexities, glitches, and limitations of machine learning.

Given the centrality of aesthetics and (in)visibility in cultivating an AI imaginary, artistic practices provide an especially fecund ground for interrogating and reworking such conventions. Joanna Zylinska further suggests that artists might even precipitate more ethical approaches to machine learning by "looking askew at the current claims and promises about AI, with their apocalyptic as well as redemptive undertone—and by retelling the dominant narratives in different genres and media."[25] I therefore turn to these artworks as important examples of both incredulous and imaginative thinking about vocal technologies. Herndon, Gee, and Audry create multiauthorial sonic assemblages that foster affects and intimacies distinct from the frictionless communication promised by voice assistants. These works are thus taken to be "*telling better stories about*

22. Mladen Dolar, *A Voice and Nothing More* (Cambridge, MA: MIT Press, 2006), 9.

23. Dolar, *A Voice and Nothing More*, 8.

24. Galloway, *The Interface Effect*, 40.

25. Joanna Zylinska, *AI Art: Machine Visions and Warped Dreams* (London: Open Humanities Press, 2020), 29–30.

AI, while also *imagining better ways of living with AI*,"[26] a practice that I suggest is necessarily intertwined with building new figurations of voice.

INVISIBLE VOICES AND SPEAKING MACHINES

The work of von Kempelen, a Hungarian inventor who debuted several mechanical curiosities in the courts of the Habsburg Empire, recurs as a touchstone both for scholars of AI and theories of voice. Von Kempelen rose to fame in 1770 with the invention of an elaborate chess-playing automaton, comprised of a life-sized wooden figure seated behind a cabinet, costumed in fur-trimmed robes and a turban. Indeed, in contrast to other popular automata of its day, the Mechanical Turk was purportedly capable not only of automated movement, but of autonomous thought.[27] Von Kempelen described his automaton as a "thinking machine," capable of deciding its own moves and masterfully executing a winning chess game on the basis of its own intelligence. The chess player was in fact an elaborate illusion controlled by a concealed human operator, yet this invention is nonetheless significant for the questions it inspired regarding the possibility of artificial intelligence as such.

The continued citation of the Mechanical Turk comprises an ongoing, if backhanded, disclosure that the concealment of human labor is integral to the functioning of seemingly intelligent machines.[28] Most prominently, in 2005 Amazon publicly launched its Mechanical Turk (MTurk) platform, which comprises a massive invisible workforce (500,000 people worldwide as of 2015[29]) that completes simple "human intelligence tasks" (HITs)—such as image tagging, audio transcribing, copywriting, data verification, and de-duplication[30]—that exceed the abilities of an algorithm. Seemingly automated operations are thus propped up by piecemeal human cognitive labor in a phenomenon that Jeff Bezos has cheekily described as "artificial artificial intelligence."[31] The platform—which reduces business clients to "requesters" rather than employers, and workers to anonymous "Turkers"—thus cultivates the expectation of inexpensive and frictionless completion of tasks that necessarily treats humans as machines.[32] Mary L. Gray and Siddarth Suri have suggested that MTurk, as one of the first commercially available platforms for crowdsourced labor, set the standards for what they term "ghost work," a veiled "digital assembly line [that] aggregates the collective input of distributed

26. Zylinska, *AI Art*, 31.

27. Elizabeth Stephens, "The mechanical Turk: a short history of 'artificial artificial intelligence,'" *Cultural Studies* 37, no. 1 (2023): 2.

28. Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (New Haven, CT: Yale University Press, 2021), 66.

29. Paul Hitlin, "Research in the Crowdsourcing Age, a Case Study," *Pew Research Center*, July 11, 2016, www.pewresearch.org/internet/wp-content/uploads/sites/9/2016/07/PI_2016.07.11_Mechanical-Turk_FINAL.pdf.

30. "Amazon Mechanical Turk," Amazon, www.mturk.com/worker.

31. Jason Pontin, "Artificial Intelligence, With Help From the Humans," *New York Times*, March 25, 2007, www.nytimes.com/2007/03/25/business/yourmoney/25Stream.html.

32. Crawford, *Atlas of AI*, 64–65.

workers."[33] The evaporation of accountability facilitated by platforms such as MTurk creates conditions of "algorithmic cruelty"[34]; indeed, a 2018 study revealed that Turkers earn a median wage of approximately $2 per hour.[35]

Yet such disclosures can hardly be considered revelations, since the very allusion to Mechanical Turk brings ghost workers into the light of day. Elizabeth Stephens draws a parallel between the kind of open secret of the Mechanical Turk and the ways in which Amazon puts forward the term "artificial artificial intelligence" as a "distractingly interesting concept" that invites a "gentle puzzlement,"[36] redirecting attention from MTurk's fundamentally exploitative business model. Observing that von Kempelen's claims regarding his thinking machine's cognitive abilities were met with public skepticism from the moment it debuted, Stephens argues that the exoticized characterization of the chess-playing figure was meant to aesthetically connote its fakery, like "a kind of magic trick whose success lay in fooling an audience aware they were being hoodwinked."[37] In this respect, Amazon's callback to the Mechanical Turk alludes not only to the integral role of concealed human labor in seemingly intelligent machines, but also to the aesthetic and political conditions of that concealment—a kind of hiding in plain sight. It is precisely this exhibition modality that is constitutive of an AI imaginary—a generative dissemblance that produces new illusions and affects.

Voice assistants, despite their promise to accomplish all manner of administrative and household tasks like magic, are also better understood as aggregates and coordinators of human labor. Vocal interfaces enlist the labor of their users to ameliorate their language processing skills, as recorded speech inputs and outputs provide ample linguistic data for machine learning.[38] Further, as is common practice in natural language processing, Alexa, Siri, and Google Assistant rely upon "thousands of low-paid humans who annotate sound snippets"[39] in order to refine their conversational abilities.[40] There are always humans in the assemblage that comprises the nonhuman speech of digital assistants, and indeed users have likely consented to participate through an end-user license agreement. Yet, as with the Mechanical Turk, such technologies vacillate between transparency and obfuscation, all in the service of ever more frictionless interfacing with digital platforms.

Further to positing speech as a purportedly natural user interface, Thao Phan suggests that the success of voice assistants requires "the perfect mimesis of the social order within

33. Mary L. Gray and Siddarth Suri, *Ghost Work: How to Stop Silicon Valley From Building a New Global Underclass* (New York: Harper Collins, 2019), ix.

34. Gray and Suri, *Ghost Work*, xxx.

35. Kotaro Hara, Abigail Adams, Kristy Milland, Saiph Savage, Chris Callison-Burch, and Jeffrey P. Bigham, "A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk," *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1.

36. Stephens, "The Mechanical Turk," 14, 15.

37. Stephens, "The Mechanical Turk," 11.

38. West, *Buy Now*, 123.

39. Austin Carr, Matt Day, Sarah Frier, and Mark Gurman, "Silicon Valley Is Listening to Your Most Intimate Moments," *Bloomberg*, December 11, 2019, bloomberg.com/news/features/2019-12-11/silicon-valley-got-millions-to-let-siri-and-alexa-listen-in.

40. West, *Buy Now*, 123.

the speech acts of the algorithm itself."[41] As digital assistants strive to facilitate frictionless encounters between users and cloud platforms, "the category of the invisible becomes, then, a performance of the socially invisible."[42] Numerous scholars have indeed suggested that the prevalent use of feminized voices in such interfaces aligns with ingrained gendered stereotypes regarding power relations in domestic and professional contexts—recalling a plucky personal assistant or submissive domestic laborer—and thereby mollify users' anxieties regarding surveillance and data mining.[43] Phan elsewhere comments upon the ways in which the utterances of digital assistants "evade specific identifying cultural inflections,"[44] while adhering to American, British, or Australian national accents (which rarely reflect actual regional specificities), thus advancing an aesthetic that conflates neutrality with a generalized whiteness. Referring specifically to Amazon, Phan suggests that all manner of tasks behind the production and functioning of Alexa are performed by predominantly racialized workers—from assembly line workers building smart devices to gig economy service workers realizing consumers' demands—and that their labor is obfuscated by the interface's white and feminized voice. The invisibility of vocal interfaces is therefore further bolstered by the mobilization of social invisibility, which is itself undergirded by the logics of whiteness and hetero-patriarchy. Thus, while the human labor integral to the functioning of voice assistants is to some degree hidden in plain sight, the stakes of these aesthetic conventions are nonetheless political.

The functionality of voice in this paradigm is both evinced and complicated by the pairing of von Kempelen's Turk with his "speaking machine," unveiled in 1783. In contrast to the illusions and trickery upon which he relied with the Mechanical Turk, the speaking machine was a meticulous attempt to replicate the acoustic productions of the human vocal apparatus. An accordion-like pump referred to as the "bellows" acted as the lungs, creating a "wind" that flowed into a "windchest" containing various mechanisms that could be manipulated with levers to produce different consonants.[45] While the mechanics within the windchest were concealed in a wooden box, the presentation of the device bore none of the anthropomorphism or exoticism of the Mechanical Turk. Von Kempelen adamantly advocated for the scientific significance of the machine, publishing a book in 1791 titled *Mechanism of Human Speech and Language* that detailed his research and experiments. However, Dolar crucially notes that the speaking machine and the Mechanical Turk were often publicly exhibited together as a kind of double bill or "double device" when von Kempelen toured across Europe in the 1780s.[46] The

41. Phan, "The Materiality of the Digital," 28.

42. Phan, "The Materiality of the Digital," 28.

43. See Heather Suzanne Woods, "Asking more of Siri and Alexa: feminine persona in service of surveillance capitalism," *Critical Studies in Media Communication* 35, no. 4 (2018): 334–49; and Amy Schiller and John McMahon, "Alexa, Alert Me When the Revolution Comes: Gender, Affect, and Labor in the Age of Home-Based Artificial Intelligence," *New Political Science* 41, no. 2 (2019): 173–91.

44. Thao Phan, "Amazon Echo and the Aesthetics of Whiteness," *Catalyst: Feminism, Theory, Technoscience* 5, no. 1 (2019): 21.

45. Leibniz Association, "The 'Kempelen' Speaking Machine," https://artsandculture.google.com/u/0/story/2QUB7hLe64FKJA?platform=hootsuite.

46. Dolar, *A Voice and Nothing More*, 8.

speaking machine was often presented first, as a kind of prelude to the thinking machine: "the former made the latter plausible, acceptable, endowed with an air of credibility."[47] Thus, in this particular origin story of artificial artificial intelligence, voice is already an integral component in sustaining the illusion of such intelligence. Jessica Riskin describes how in eighteenth-century debates regarding the possibility and limitations of mechanical imitations of life, spoken language, along with perpetual motion, was situated "at the crux of the distinction between animate and inanimate, human and nonhuman."[48] While the Turk itself never spoke, it performed on the epistemological stage established by the demonstration of mechanical speech.

The fact that machine intelligence appeared plausible when coupled with the innately human characteristic of speech attests to the power of the vocal imaginary, as if the agential properties of voice could be transferred to the Mechanical Turk by proximity. Yet this demonstration also unsettled these very distinctions, defying dominant beliefs that voice was too organic a process ever to be simulated. This ambivalence inherent to von Kempelen's double device is proffered as an exemplary anecdote for Dolar's theory of voice—a framework that problematizes the emphasis on innateness and invisibility outlined above. Despite the transparency of the material components that together generate the speaking machine's utterance, and their alignment with the familiar elements of the human speech apparatus, the device was nonetheless received by the public as an eerie enigma. As Dolar describes,

> There is an uncanniness in the gap which enables a machine, by purely mechanical means, to produce something so uniquely human as voice and speech. It is as if the effect could emancipate itself from its mechanical origin, and start functioning as a surplus—indeed, as the ghost in the machine; as if there were an effect without a proper cause, an effect surpassing its explicable cause.[49]

This observation regarding the strangeness of the nonhuman voice crucially exposes, for Dolar, a vocal topology shared with the human voice. In every instance, voice is always irreducible to the means of its production, whether by the fleshy apparatus of the lungs and larynx or other mechanical means.[50] The very fact that voice can be produced by machines situates it in "a zone of undecidability, of a between-the-two, an intermediacy," which marks "one of the paramount features of the voice."[51] The relationship between synthesized speech and thinking machines is thus, in this telling, more convoluted than it initially appears: more than a design feature that obfuscates, naturalizes, and lends credibility to the latter, this presentation of voice generates unintended and uncanny affects that inspire a renewed consideration of the metaphysics of voice as such.

47. Dolar, *A Voice and Nothing More*, 9.

48. Jessica Riskin, "The Defecating Duck, or, the Ambiguous Origins of Artificial Life," *Critical Inquiry* 29, no. 4 (Summer 2003): 617.

49. Dolar, *A Voice and Nothing More,* 7–8.

50. Dolar, *A Voice and Nothing More*, 70.

51. Dolar, *A Voice and Nothing More*, 13.

Similar, perhaps, to the ways in which voice is deployed in smart technology as an invisible interface, Dolar's formulation describes voice as a "vanishing mediator,"[52] a material support that disappears in the meaning that emerges through it. Yet voice simultaneously refuses the reduction to meaning, always leaving audible remainders such as timbre, accent, and intonation—what Dolar calls an "excrement of the signifier."[53] Further, it is precisely the purging of these extra-signifying elements in mechanical voices that paradoxically allow its "disturbing and uncanny nature"[54] to emerge. Voice can therefore never quite vanish; it operates in remainders and excesses that unsettle the perception of speech as frictionless communication. As an interface, it is always unworkable. While the selection of white, feminine, or otherwise "socially invisible" voices might be understood as an attempt to suture the uncanny valley that opens up as humanoid machines approach realism,[55] it is crucial that such affects remain and indeed flourish. Indeed, popular counter-discourses highlight communicative glitches and anomalous utterances produced by voice assistants[56] such as spontaneous outbursts of laughter.[57]

I do not wish to install Dolar's metaphysical claims as a more accurate or entirely unproblematic way to think about voices,[58] but his perspective does offer an intriguing point of entry to consider these affective perturbations and their implications for a vocal imaginary as it is wedded to an AI imaginary. Rather than considering synthesized voices as impoverished or falsified renderings of a "real" voice, this framework places the vocalizations of human and nonhuman actors in tandem, thereby rattling the ideological foundations that exceptionalize the voice as a privileged and innately human modality. The artworks discussed in the last section of this paper take up this invitation and further explore the possibilities afforded through a consideration of voice, not as a stable, singular entity, but as a transindividual process.[59]

"...better stories about AI"

Having delineated the ideological foundations of the current sociotechnical milieu, and its historical antecedents, the final section of this essay examines how several artworks—Herndon's *PROTO* and a series by Audry and Gee comprising *of the soone, to the sooe*, and *Machine*

52. Dolar, *A Voice and Nothing More*, 15.

53. Dolar, *A Voice and Nothing More*, 20.

54. Dolar, *A Voice and Nothing More*, 22.

55. AO Roberts, "Echo and the Chorus of Female Machines," *Sounding Out!* (blog), March 2, 2015, https://soundstudiesblog.com/2015/03/02/echo-and-the-chorus-of-female-machines.

56. See, for instance, Paul Lamkin, "The creepiest, freakiest things Alexa has ever said and done," *Ambient*, February 11, 2022, www.the-ambient.com/features/the-creepiest-freakiest-things-alexa-has-ever-said-and-done-2154; Katie Teague, "10 weirdest things Alexa can do on your Amazon Echo," *CNET*, September 20, 2020, www.cnet.com/home/smart-home/10-weirdest-things-alexa-can-say-and-do-on-your-amazon-echo.

57. Julia Carrie Wong, "Amazon working to fix Alexa after users report random burst of 'creepy' laughter," *Guardian*, March 7, 2018, www.theguardian.com/technology/2018/mar/07/amazon-alexa-random-creepy-laughter-company-fixing.

58. For a critique of Dolar's perspective see Mickey Vallee, "Possibility, Performance, Politics: On the Voice and Transformation," *Parallax* 23, no. 3 (2017): 330–41.

59. Rachelle Chadwick, "Theorizing voice: toward working *otherwise* with voices," *Qualitative Research* 21, no. 1 (February 2021): 79.

Unlearning—have challenged the conventions and imaginaries perpetuated in and by vocal interfaces. It is crucial, however, to observe that the realm of artistic production is by no means exempt from the practices of obfuscation that constitute the AI imaginary outlined above. Zylinska observes that much generative art takes a celebratory stance, creating striking visuals to showcase and spectacularize the volume of data so speedily processed by AI systems, and is indeed often facilitated or sponsored by a major platform such a Google, Amazon, or Apple.[60] Citing works such as Mario Klingemann's *Superficial Beauty* (2017) or *Neural Face* (2017), in which the artist has trained generative adversarial networks (GANs) to produce photorealistic portraits of "new" faces based on datasets scraped from large online repositories, Zylinska suggests that such works advance a new aesthetic that is "slightly uncanny, already boring."[61] Yet what is more problematic about this genre of artistic engagement with AI is that it "is premised on the banality of looking, with perception understood as visual consumption and art reduced to mild bemusement."[62] Such art practices thus participate in an exhibition culture similar to that of the Mechanical Turk; although less preoccupied, perhaps, with trickery, these generative artworks mobilize an alien aesthetic—albeit one of the digital sublime rather than orientalist other—to mobilize public curiosity and credulity, rather than criticality, regarding the potentials of AI.

Works such as Klingemann's, and indeed recent image-generating applications such as DALL-E and Lensa AI,[63] suggest that the capabilities of machine learning can be affirmed according to a paradigm of fidelity. The GAN's outputs are original, yet their success is measured according to how convincingly they emulate or interpret their visual or textual inputs. Such applications rely on the availability of large datasets to train algorithms to generate photorealistic figures or imitations of canonical artistic styles. Indeed, so too would any artistic exploration seeking to generate a voice of a similar fidelity or "high modality"[64]—a requirement that restricts such a possibility to major corporate actors. I suggest, however, that Audry, Gee, and Herndon, working with smaller, idiosyncratic, and intimate datasets, advance different aspirations in their practices. Rather than seeking to wow spectators with the magic verisimilitude of their vocal creations, they amplify the messiness and mistranslations within the hybrid human/nonhuman processes and interactions that generate them.

The art practices under discussion might be therefore understood as engagements with varied vocal interfaces that make audible the intrafaces, the zones of indecision, internal to them. Although the final works bear little resemblance to voice assistants, nor do they

60. Zylinska, *AI Art*, 75–76.

61. Zylinska, *AI Art*, 81.

62. Zylinska, *AI Art*, 81.

63. Although a thorough discussion exceeds the scope of this paper, it bears noting that critical perspective toward DALL-E and Lensa AI observe that these applications rely on enormous datasets comprised of copyrighted images, essentially stealing from previously existing artwork. Further scholarship might consider how this process renders artists as unconsenting ghost workers. See Rhea Nayyar, "Read This Before You Jump on the Lensa 'Magic Avatar' Trend," *Hyperallergic*, December 6, 2022, https://hyperallergic.com/785759/read-this-before-you-jump-on-the-lensa-magic-avatar-trend.

64. Justine Humphry and Chris Chesher, "Preparing for smart voice assistants: Cultural histories and media innovations," *New Media & Society* 23, no. 7 (2020): 1983.

utilize precisely the same natural language models, the vocal imaginary that endows a sense of naturalness or credibility to thinking machines provides the sociotechnical backdrop from which these practices emerge. Indeed, Gee situates "cheerful, often female-gendered voices of personal assistants"[65] as the ground she wishes to interrogate. As Galloway notes, interfaces are always unworkable, but never admit their own unworkability.[66] It is precisely this unworkability of vocal interfaces that Audry, Gee, and Herndon seek to tune into by unravelling the vocal imaginary that sustains the illusion of their neutrality, invisibility, or self-sufficiency.

The series of collaborative artworks by Gee and Audry demonstrates an irreverent approach to the myths of impartiality and intelligence conjured by the contemporary AI imaginary. Audry designed a deep-learning neural network, which was then trained using text from Emily Brontë's 1847 novel *Wuthering Heights*. The algorithm read and reread the text letter by letter, rather than word by word, gaining familiarity with the "syntactical universe"[67] of Brontë's writing before generating its own text seeking to emulate her authorial voice. The result documents the algorithm's learning process, as it first generates a stuttering collection of letters and phonemes punctuated by gaps and repetitions:

> . . . e e h a h a o h o o a a o o a ao i a o o o a h a te aa a ae aee
>
> ae ae a a a he he a ah a ae he h te a te ae io ae ao aa ao ao ao ao ao ao ao ao ao ao aoe ao ao ao ao ao ao aa ao ao ao ao ie a a te a a a tea o aa ao a a a aa a te a a ao a an ao a te ao a ao an te to ao ao ao an te ao a he ie a io aw ao ao ah ao ao ao a tee an an a tee tee the soe an an an tee and an an and an tee an an aa tee an te tee . . .[68]

The neural net's output slowly evolves, piecing together recognizable words and turns of phrase. The text begins to conform more closely to the vocabulary and style that might characterize a nineteenth-century gothic novel, yet the algorithm remains unable to articulate defined sentences, let alone a cohesive narrative:

> . . . i was a sente of the soow of a second minutes, and the sense of her hands and said and said the master was all the hearth, and the same silence of the hand of the master then the master was the staying and sente of the same stare of a servants of the house of the front shoulders to the servants of the hearth was she was a servant to her father, and she was a little and seen to the window . . .[69]

Audry and Gee thus present AI as in process and imperfect—a notable contrast to the coherent and already smart personas advanced by voice assistants. In further contrast to the professed neutrality of machine learning, Audry and Gee stress the degree to which "the aesthetics of this textual output are heavily determined by Audry as a human artist-programmer."[70] The selection of Brontë's stylized text as a training set makes apparent

65. Erin Gee and Sofian Audry, "Automation as Echo," *ASAP/Journal* 4, no. 2 (2019): 307.
66. Galloway, *The Interface Effect*, 52–53.
67. Gee and Audry, "Automation as Echo," 308.
68. Gee and Audry, "Automation as Echo," 309.
69. Gee and Audry, "Automation as Echo," 309.
70. Gee and Audry, "Automation as Echo," 308.

to the sooe (2018) by Sofian Audry and Erin Gee; photograph by Hexagram; copyright Erin Gee 2018.

the degree to which such datasets, and the humans who create and define them, shape the epistemic boundaries of what AI can do, say, or recognize.

The three resulting artworks represent different attempts to interpret the generated output as script or score, with Gee vocalizing the fragmented and nonsensical text using

techniques associated with autonomous sensory meridian response (ASMR), a relaxing, euphoric sensory response to audiovisual triggers, often referred to as "tingles," that provides the basis for an online community that creates and consumes videos seeking to trigger such a physiological response.[71] ASMR artists and influencers have introduced a distinctive vocabulary of performative conventions such as tapping, brushing, and scratching everyday objects, as well as gentle whispering and other extralinguistic oral sounds (including breathing, tongue clicking, and chewing), which Gee incorporates into her sound art practice. The first piece using the AI-generated score, an audio work titled *of the soone,* was recorded using amateur equipment in Gee's kitchen.[72] This was followed by *to the sooe,* which comprises a 3D-printed sculpture etched with text generated by the neural network, as well as audio of Gee's vocal interpretation, this time recorded using advanced ambisonic recording methods during a residency at the Institute for Electronic Music and Acoustics in Graz, Austria. Finally, *Machine Unlearning* comprises a roleplay video in which Gee embeds another performance of the generated score within the dramaturgical conventions of ASMR. Gee addresses the viewer directly, positing her vocal performance as a kind of "treatment" before she recites the score backward, as if the neural net is undoing its training.

Machine Unlearning, *of the soone*, and *to the sooe* generate frictions that belie the ideology of vocal frictionless-ness upon which the AI imaginary relies. Gee describes the works as the product of three "authorial agencies": Brontë, Gee as an ASMR interpreter, and the neural network,[73] in addition, of course, to Audry as its programmer. The compositions are produced as a collaboration between human and nonhuman actors, irreducible to a singular author. Voice becomes a tool to sound out this aggregation, rather than to lend credence to the abilities of a miraculous thinking machine. With this in mind, a comparison might be drawn between Gee's virtuosic performance and those undertaken by actors who lend their voices to interfaces such as Alexa, Siri, and Google Assistant. Although the identities of these human performers are rarely disclosed, Susan Bennett, the voice of the original Siri, went public in 2013, describing how she recorded hours of nonsensical phrases such that phonemes could be isolated and extracted by developers and algorithmically stitched back together.[74] Around the same time, voice actor September Day described reciting passages from *Alice in Wonderland,* the AP news wire, and miscellaneous numbers in different cadences for six to seven hours a day for eight days in order create the text-to-speech application incorporated into Amazon's Kindle Fire tablet.[75] While I do not mean to suggest that this labor undertaken predominantly by a select few white women is subject to the same degree of exploitation as

71. Naomi Smith and Anne-Marie Snider, "ASMR, affect and digitally-mediated intimacy," *Emotion, Space and Society* 30 (2019): 41.

72. Erin Gee, email message to author, June 21, 2022.

73. Erin Gee, "to the sooe," https://eringee.net/to-the-sooe.

74. Jessica Ravitz, "'I'm the original voice of Siri,'" *CNN*, October 15, 2013, www.cnn.com/2013/10/04/tech/mobile/bennett-siri-iphone-voice/index.html.

75. Lessley Anderson, "Machine Language: How Siri Found Its Voice," *The Verge*, September 17, 2013, www.theverge.com/2013/9/17/4596374/machine-language-how-siri-found-its-voice.

Turkers or other ghost workers, it nonetheless comprises an often-overlooked human element integral to the functioning of vocal interfaces. Indeed, these invisible vocal performances are applied in service of the interface's invisibility. By bringing to the fore Gee's vocal performance that (perhaps inadvertently) resembles the peculiar human performances that generate the sonic material for natural language synthesis, Gee and Audry reshuffle the relational networks that comprise generated voices.

The use of vocal techniques affiliated with ASMR also contributes to an aesthetic reorientation of the vocal imaginary, one that embraces extra-communicative and affective dimensions of voice advanced via Dolar's theory of voice. ASMR creates, according to Gee and Audry, "a slow bath of evolving sound [rather] than a clear or graspable process"[76]—a shifting zone of affective encounters rather than a tidy data input/output or command/response exchange. All of the audible elements extraneous to speech—to borrow Dolar's phrase, the "excrement of the signifier"—flourish in this context, valorized over the speed, clarity, and efficiency of information exchange. In this respect, ASMR is "almost content free," preoccupied with making accessible triggers that generate affective experiences.[77]

Moreover, I suggest that as a mode of vocal performance developed with digital tools and for dissemination via digital networks, ASMR presents a genre of vocalization that is markedly transindividual—collectively formed and reformed in constant relation to a sociotechnical milieu. Although ASMR works to activate physiological responses in the viewer or listener and evokes the somatic presence of the performer through bodily sounds like breathing and swallowing, the prominence of the human body ought not be interpreted as a recourse to an organic or originary vocal relationality. To the contrary, ASMR constitutes a highly mediated performance in every respect: it is facilitated by amplification devices to pick up on all of the sonic minutiae of the artists' haptic and extralinguistic gestures, requires listeners to wear headphones to achieve an immersive binaural effect, and is distributed and accessed via online networks. It is "a digitally-mediated affective experience uniquely shaped by online spaces and their affordances."[78] This resonates with the understanding of all voices advanced by Rachelle Chadwick—as entanglements between bodies, machines, and sociocultural relations—as a counter to the romantic humanism with which they are so often endowed. Voices are not stable, authentic, and singular but rather "sites in which the radical permeability between bodies, ideologies, selves, sociocultural relations, machines and biologies are enacted."[79] Considered in this way, voice is too porous, too enmeshed for the ideological work it is meant to perform with regard to smart technologies. How can voice lend credence to AI's stable and unbiased rationality, if voice itself is contingent and processual?

Artist and composer Herndon also embraces a transindividual understanding of voice on her album *PROTO*, which was created in collaboration with a voice-processing neural

76. Gee and Audry, "Automation as Echo," 310.
77. Smith and Snider, "ASMR, affect and digitally-mediated intimacy," 43.
78. Smith and Snider, "ASMR, affect and digitally-mediated intimacy," 41.
79. Chadwick, "Theorizing Voice," 91.

Vision calibration (still) from *Machine Unlearning* (2018–19) by Sofian Audry and Erin Gee; photograph by Elody Libe; copyright Erin Gee 2020.

network built with her partner/fellow artist Mat Dryhurst and developer Jules LaPlace. While Audry and Gee trained their neural net using an existing textual input and then relied on Gee's embodied voice "as a relatively low-tech filter for processes of machine automation,"[80] Herndon's process involved creating an original piece of music, which she then recorded or taught to a choral ensemble. Her composition then acted as a dataset to train the neural net using sonic inputs, which in turn generated its own audible outputs.[81] Over the course of creating the album, a group of vocalists met to "feed" Herndon's custom AI housed in a gaming PC, which she called "Spawn" and assigned she/her pronouns. The members of the choral ensemble sang and talked to Spawn, who likewise sang and spoke back. Herndon concedes that the process of training Spawn was arduous, remarking that "everything sounded like ass" for the first six months of experimentation.[82] The final results as they are heard on *PROTO* are eclectic—buzzing choral melodies, thunderous clashing beats, and indecipherable spoken-word samples from human and nonhuman collaborators alike.

Although Herndon's work hardly resembles the conversational tone or linguistic clarity of a voice assistant, a comparison might be made between the ways that they operate, at their most basic level, as voice-processing interfaces. Voice assistants convert users' utterances into executable commands and respond to them, as well as capture these utterances as training data to further advance their vocal capabilities. Spawn likewise listens, responds, and cultivates her voice, albeit using sonic data inputs from a small group of consenting and compensated professional singers and musicians rather than a vast number of mostly unwitting consumers of smart speakers. Further, by attending to musical parameters and audible vocal traits rather than linguistic meaning, Spawn exceeds the command/response model that dominates interactions with digital assistants[83] and generates unexpected outputs. Herndon takes up a malleable approach to working with machine learning; Spawn acts as a compositional tool, a musical instrument, and a performer within a broader ensemble, with these roles shifting and recombining in different ways on different tracks on *PROTO*. "Canaan" and "Evening Shades" are both described as "live training" sessions; the former a lyrical a cappella performance by three singers, including Herndon, offering up their voices as data, and the latter a call-and-response between a large choral ensemble and Spawn, who sings back in a hissing echo. Whereas for "Godmother," Spawn was trained with percussion tracks by footwork electronic musician and producer Jlin and generates her own stuttering beats using Herndon's voice.

80. Gee, "to the sooe."

81. For additional details on this process, see Katie Hawthorne, "Holly Herndon: the musician who birthed an AI baby," *Guardian*, May 2, 2019, www.theguardian.com/music/2019/may/02/holly-herndon-on-her-musical-baby-spawn-i-wanted-to-find-a-new-sound; and Holly Herndon, "Holly Herndon–AI is not going to kill us; it might make us more human," interview by Stuart Stubbs, *Loud and Quiet,* April 30, 2019, www.loudandquiet.com/interview/holly-herndon.

82. Herndon, "Holly Herndon–AI is not going to kill us."

83. Simone Natale and Henry Cooke, "Browsing with Alexa: Interrogating the impact of voice assistants as web interfaces," *Media, Culture & Society* 43, no. 6 (2021): 1007.

Cover of Holly Herndon's 2019 album *PROTO*; cover design by Michael Oswell for 4AD.

In several respects, Herndon approaches, yet convolutes, the metaphors and conventions that characterize popular understandings of AI. The anthropomorphizing of Spawn certainly resonates with the ways in which vocal interfaces are also assigned feminized personas. Yet, this figuration is crucially different as she refers to Spawn as her "inhuman child"[84] and metaphors of AI babies abound in reviews and interviews surrounding the release of *PROTO*. Unlike Siri, Alexa, or Google Assistant, who present themselves as fully formed, already "smart," and ready at their users' beck and call, Spawn more readily reveals the constant care and human attention that such systems demand.[85] She "requires a community to raise her"[86] and is entirely susceptible to the aesthetic intentions, inputs, and biases of that human community. Insofar as the voices heard on *PROTO* comprise

84. Holly Herndon, "Inhuman After All," interview by Gabriela Tully Claymore, May 6, 2019, www.stereogum.com/2041686/holly-herndon-proto-interview/interviews.

85. Gray and Suri, *Ghost Work*, xviii.

86. Herndon, "Inhuman After All."

"a weird hybrid ensemble where people are singing with models of themselves,"[87] Spawn might also be interpreted in relation to "statistical doubles"[88] of human consumers, which are generated through algorithmic data capture and designed to predict our desires and purchasing patterns. For Schmuckli, such doubles reside in what she calls a "reconfigured uncanny valley"[89] that exceeds the feeling of disorientation caused by humanoid automata and is instead "mapped by the inscrutable calculations of algorithms that are designed to mine and analyze humans' behavior and project it into tradable futures."[90] Indeed, Spawn holds up a vocal mirror to her trainers, singing their own voices back with a difference.[91] This dynamic is clearly audible in "Evening Shades," where the chorus sings and Spawn sings back, dropping notes and garbling lyrics. Spawn is incapable of perfectly recreating the complexity of human voices, just as statistical doubles can never entirely reflect one's complete self, motivations, and desires—although they might get eerily close. *PROTO* might therefore be understood as a dwelling within this uncanny doubling, reveling in the gaps and discrepancies between Spawn and her trainers, as much as the resemblances.

In this respect, Herndon's approach is also distinguished from the aforementioned approaches to generative art that consider fidelity as an affirmation of intelligence. Indeed, Herndon is critical of applications that involve statistical analysis of musical scores in order to emulate an artist or style, such as Amper or Jukebox; this constitutes "an aesthetic cul-de-sac"[92] in which the measure of AI's "creativity" is its ability to mimic that which already exists. Instead, Herndon's emphasis upon Spawn's role as part of a collaborative musical ensemble seems to draw an unexpected parallel, perhaps, between training an algorithm and the vocal training undertaken by human singers and choruses. The effect is not to endow Spawn with a kind of elevated or anthropomorphic agency associated with human voices, but rather to highlight that all musical expressions of voice are necessarily technological. Indeed, Western vocal pedagogy, which is crucially informed by medical measuring and imaging practices such as laryngoscopy, advances an understanding of voice as an instrument.[93] A somewhat paradoxical presumption emerges from this discourse: that one's authentic voice can only be "discovered" through instruction,

87. "Holly Herndon on the power of machine learning and developing her 'digital twin' Holly+," interview by Jordan Darville, *The FADER*, July 27, 2021, www.thefader.com/2021/07/27/holly-herndon-on-the-power-of-machine-learning-and-developing-her-digital-twin-holly.

88. Schmuckli, "Automatic Writing and Statistical Montage," 15.

89. Schmuckli, "Automatic Writing and Statistical Montage," 9.

90. Schmuckli, "Automatic Writing and Statistical Montage," 15.

91. Following *PROTO,* Herndon's project *Holly+* (2021–present) offers a higher fidelity model of her voice to other creators, inviting them to make original works with Herndon's digital likeness. This work—which disburses Herndon's voice yet affirms her proprietary relationship to it—presents a different approach to vocal imaginaries and AI than that explored in her earlier work. While discussion of *Holly+* exceeds the scope of this paper, it merits further consideration. See Holly Herndon, "Holly+," https://holly.plus.

92. Holly Herndon, "Holly Herndon on Her AI Baby, Reanimating Tupac, and Extracting Voices," interview by Emily McDermott, *Art in America*, January 7, 2020, www.artnews.com/art-in-america/interviews/holly-herndon-emily-mcdermott-spawn-ai-1202674301.

93. Sterne, *Diminished Faculties*, 97.

Holly Herndon and her collaborators, including Spawn on top of the piano; photograph by Boris Camaca.

rehearsal, and exercise.[94] As Herndon remarks, "the voice isn't necessarily individual; it belongs to a community, to a culture, to a society."[95] Spawn's voice, like those of her human trainers and fellow performers, is ever becoming entangled with all manner of material and discursive objects and relations. *PROTO* is thus also emblematic of a trans-individual approach to voice—a "process that happens between bodies, locations, affective and discursive histories."[96]

CONCLUSION

Herndon states that "AI is just us. AI is human labor obfuscated through a terminology called AI."[97] This paper has argued that voice, and the ideological convictions that surround it, play a crucial role in this obfuscation that is central to the very functioning of AI. Smart devices rely upon an imaginary predicated upon the naturalness of speech—mobilized alongside mythical whiteness and femininity—to render vocal interfaces, and the human labor that supports them, invisible. Voice is situated as a frictionless, mediation-less medium, one that lends credence to the seamless integration, intelligent capacities, utility, and neutrality of algorithms more broadly. Therefore, artistic practices

94. Katherine Meizel, *Multivocality: Singing on the Borders of Identity*(Oxford: Oxford University Press, 2020), 25–28.

95. Herndon, "Holly Herndon on Her AI Baby."

96. Chadwick, "Theorizing Voice," 91.

97. Herndon, "Holly Herndon on Her AI Baby."

that situate voice in varied and reciprocal relations to machine learning offer fecund ground for thinking with and beyond the contemporary convergence of vocal and AI imaginaries. Audry, Gee, and Herndon crucially drain, revise, and reorient the obfuscating capacities and humanist ideologies at work in voice assistants. Rather than mobilizing voice as a means to render interfaces transparent, these vocal creations amplify frictions within this zone of indecision. These artistic practices therefore problematize the narratives of tectonic paradigm shifts and sensory revolutions laid out at the beginning of this paper. By presenting voice as transindividual process—the property of neither human subjects nor machines, but as relational sites that emerge between them—these artworks unsettle the ideological foundations upon which such ebullient stories about voice, and indeed about AI, rely, thus creating an opening for thinking otherwise.

This paper also has sought to demonstrate the possibilities afforded by bringing together insights from sound and voice studies with critical perspectives on AI. Such scholarship increasingly attends to the ways in which voices are "sites where laryngeal structures intersect with power structures,"[98] prior to and as they become enmeshed with technologies of transmission, amplification, recording, and synthesis. As Amanda Weidman proposes, "the production of voices at least partly through nonhuman sources . . . can be explored for the kinds of culturally specific ideologies of intimacy, sincerity, and authenticity they engender, the aesthetics of voice with which they operate, and the new forms of subjectivity to which they give rise."[99] This is precisely the genre of exploration I have endeavored to sketch here, as such imaginaries are cultivated, and indeed lived and felt, in tandem with popular understandings and applications of AI. At a time when artists are "blowing some much-needed cool air on the heat and hype around AI currently emitting from tech companies,"[100] an interrogation of the ascendant vocal aesthetics in smart technology, and the imaginaries upon which they both rely and engender, is of vital importance to this project. ■

Alex Borkowski is a PhD candidate in the Joint Graduate Program in Communication & Culture at York University and Toronto Metropolitan University.

98. Meizel, *Multivocality*, 15.
99. Amanda Weidman, "Anthropology and Voice," *Annual Review of Anthropology* 43 (2014): 41.
100. Zylinska, *AI Art*, 31.