Voice as Ecology: Voice Donation, Materiality, Identity

Steph Ceraso

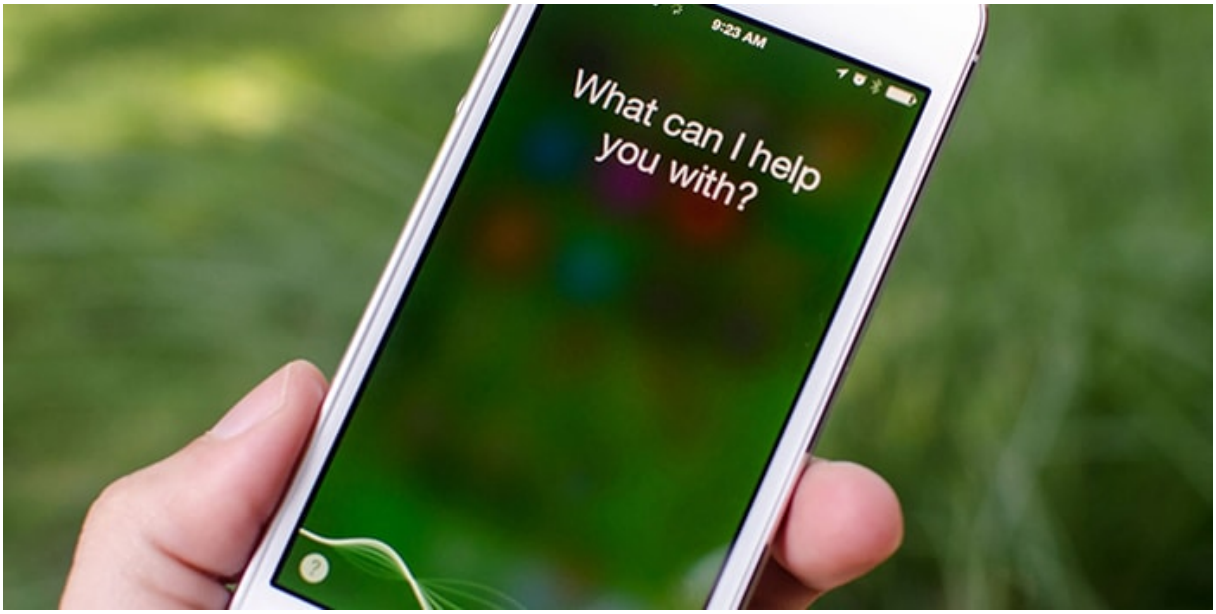
Blog post published September 6, 2022 in *Sounding Out!*
https://doi.org/10.59350/ehj00-cem02

Abstract

I first heard about voice donation while listening to "Being Siri," an experimental audio piece about Erin Anderson donating her voice to Boston-based voice donation company, VocaliD. Like a digital blood bank of sorts, VocaliD provides a platform for donating one's voice via digital audio recordings.

Keywords

Music (Subfield), Digital Media, Listening, Performance, Sampling And Remix Culture, Sound Studies



Recommended citation

Ceraso, S. (2022, September 6). Voice as Ecology: Voice Donation, Materiality, Identity. *Sounding Out!*. https://doi.org/10.59350/ehj00-cem02

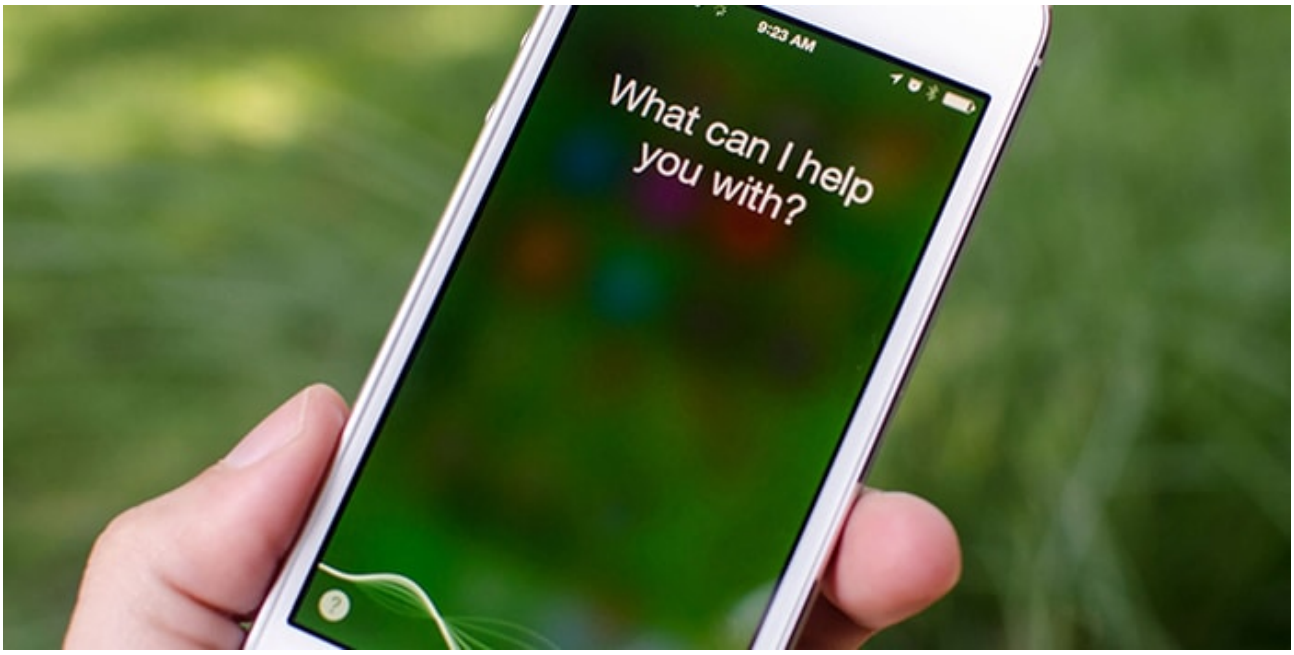
I first heard about voice donation while listening to "Being Siri," an experimental audio piece about Erin Anderson donating her voice to Boston-based voice donation company, VocaliD. Like a digital blood bank of sorts, VocaliD provides a platform for donating one's voice via digital audio recordings. These recordings are used to help technicians create a custom digital voice for a voiceless individual, providing an alternative to the predominately white, male, mechanical-sounding assistive technologies used by people who cannot vocalize for themselves (think Stephen Hawking). VocaliD manufactures voices that better match a person's race, gender, ethnicity, age, and unique personality. To me, VocaliD encapsulates the promise, complexity, and problematic nature of our current speech AI landscape and serves as an example of why we need to think critically about sound technologies, even when they appear to be wholly beneficial.

Given the extreme lack of sonic diversity in vocal assistive technologies, VocaliD provides a critically important service. But a closer look at both the rhetoric used by the organization and the material process involved in voice donation also amplifies the limits of overly simplistic, human-centric conceptions of voice. For instance, VocaliD rhetorically frames their service by persistently linking voice to humanity—to self, authenticity, individuality. Consider the following statements made by Rupal Patel, CEO and founder of VocaliD, in which she emphasizes the need for voice donation technology:

> *"Here's a way for us to acknowledge these individuals as unique human beings." (Fast Company)*
>
> *"I was talking to [a] girl we made a voice for. She told me that people are finally seeing her for who she really is." (Medieros)*

These are just a few examples from a larger discourse that reinforces the connection between voice and humanity. VocaliD's repeated claims that their unique vocal identities humanize individuals imply that one is not fully human unless one's voice *sounds* human. This rhetoric positions voiceless individuals as less than human (at least until they pay for a customized human-sounding voice).

VocaliD's conflation of voice and humanity makes me wonder about the meaning of "human" in this context. For example, notions of humanity have been historically associated with Western whiteness—and deployed as a means of separating or distinguishing white people from Others—as Alexander Weheliye points out. Though VocaliD's mission is to diversify manufactured voices, is a "human-sounding" voice still construed as a white voice? Does sounding human mean sounding white? Even if there is a bank of sonically diverse voices to choose from, does racial bias show up in the pacing, phrasing, or inflection caused by the vocal technology?



Photocredit: iphonedigital @Flickr CC BY-SA

I am also disturbed by the rhetoric of humanity and individuality used by VocaliD because the company adopts the same rhetoric to describe the AI voices they sell to brands for media and smart products. Here's an example of this rhetoric from the VocaliD AI website: "When you need a voice that resonates, evokes audience empathy, and sounds ***like you***, rather than your competitors, VocaliD's AI-powered vocal persona is the solution. ***Your*** voice — always on, where you need it when you need it." Using similar rhetorical strategies to describe both voiceless people and products is *de*humanizing. And yet, having a more diverse AI vocal mediascape, especially in terms of race, is crucially important since voice-activated machines and products are designed largely by white men who end up reinforcing the sonic color line.

Interestingly, the processes VocaliD uses to create a custom voice reveal that these voices are not, in fact, unique markers of humanity or individuality. It's hard to find a detailed account of how VocaliD voices are made due to the company's patents, but here are the basics: VocaliD does not transfer a donated voice directly to a voiceless person's assistive technology. VocaliD technicians instead blend and digitally manipulate the donated voice with recordings of the noises a voiceless person can make (a laugh, a hum) to create a distinct new voice for the recipient. In other words, donated voices are skillful remixes that wouldn't be possible without extracting vocal data and manipulating it with digital tools. Despite perpetuating narratives about voice, humanity, and authenticity, VocaliD's creative blending of vocal material reveals that donated voices are the result of compositional processes that involve much more than people.

Further, considering VocaliD voices from a material rather than human-centric perspective amplifies something important about voices in general. All voices are composed of and grounded in an ecology. That is, voices emerge and are developed through a mixture of: (1) biological makeup (or technological makeup in the case of machines with voices); (2) specific environments and contexts (geography may determine the kind of accents humans have; AI voices have distinct sounds for their brands); (3) technologies (phones, computers, digital recorders and editors, software, and assistive technologies preserve, circulate, and amplify voices); and (4) others

(humans often emulate the vocal patterns of the people they interact with most; many machine voices also sound like other machine voices). Put simply, all voices are intentionally and unintentionally composed over time—shaped by ever-changing bodily (and/or technological) states and engagements with the world. Voices are dynamic compositions by nature. Examining voice from a material standpoint shows that voices are not static markers of humanity; voices are responsive and malleable because they are the result of a complex ecology that involves much more than a "unique" human being.

However, focusing solely on the material aspects of vocality leaves out people's lived experiences of voice. And based on online videos of VocaliD recipients—like Delaney, a seventeen-year-old with cerebral palsy—VocaliD voices seem to live up to the company's hype. Delaney appears delighted by her new voice, stating: "I was so excited to get my own voice. I used to have a computer voice and now I sound like a girl. I like that. And I talk more." Delaney's teachers also discuss how her new voice completely changed her demeanor. Whereas before Delaney was reluctant to use her assistive technology to speak, her new voice gives her confidence and a stronger sense of identity. As her teacher explains in the video, "she is really engaged in groups, she wants to share her answers, she's excited to talk with friends. It's been really nice to see." For Delaney, a VocaliD voice represents a newfound sense of agency.

It's important to recognize this video is not necessarily representative of every VocaliD recipient's experience, or even Delaney's full experience. As Meryl Alper notes in Giving Voice, these types of news stories "portray technology as allowing individuals to 'overcome' their disability as an individual limitation, and are intended to be uplifting and inspirational for able-bodied audiences" (27). While we should be wary of the technological determinism in the video, observing Delaney use her VocaliD voice—and listening to the emotional responses of her mom and teachers—makes it difficult to deny that donated voices make a positive impact. For me, this video also gets at a larger truth about humans and voice: the ways we hear and understand our own voices, and the ways others interpret the sounds of our voices, matter a great deal. Voices are integral to our identities—to the ways we understand and think about ourselves and others—and the sounds of our voices have social and material consequences, as the SO! Gendered Voices Forum illustrates so clearly.

An image VocaliD used to advertise themselves on Twitter. Image used for purposes of critique.

It's worth repeating that VocaliD's mission to diversify synthetic voices is incredibly important, especially given the restrictive vocal options available to voiceless individuals. It's also necessary to acknowledge the company has limitations that end up reproducing the structural inequities it tries to address. As Alper observes, "In order to become a speech donor, one must have three to four hours of spare time to record their speech, access to a steady and strong Internet connection, and a quiet location in which to record" (162-63). With these obstacles to donating one's voice in mind, it's not surprising that all the VocaliD recipient videos I could find feature white people. Donating one's voice is much easier for middle to upper class white people who have access to privacy, Internet, and leisure time.

This brief examination of VocaliD raises questions about what a more equitable future for vocal technologies might look/sound like. Though I don't have *the* answer, I believe that to understand the fullness of voice, we can't look at it from a single perspective. We need to account for the entire vocal ecology: the material (biological, technological, financial, etc.) conditions from which a voice emerges or is performed, *and* individual speakers' understanding of their culture, race, ethnicity, gender, class, ability, sexuality, etc. An ecological approach to voice involves collaborating with people and their vocal needs and desires—something VocaliD models already.

But it also involves accounting for material realities: How might we make the barriers preventing a more diverse voice ecosystem less difficult to navigate—especially for underrepresented groups? In short, we must treat voice holistically. Voices are more than people, more than technologies, more than contexts, more than sounds. Understanding voice means acknowledging the interconnectedness of these things and how that interconnectedness enables or precludes vocal possibilities.

—

Featured image: *366-350 You can't shut me up*, Jennifer Moo, CC BY-ND

—

Steph Ceraso *is an associate professor of digital writing and rhetoric at the University of Virginia. Her 2018 book, Sounding Composition: Multimodal Pedagogies for Embodied Listening, proposes an expansive approach to teaching with sound in the composition classroom. She also published a digital book in 2019 called Sound Never Tasted So Good: 'Teaching' Sensory Rhetorics —an exploration of writing, sound, rhetoric, and food. She is currently working on a book project that examines sonic forms of invention in various contexts.*

—

REWIND! . . .If you liked this post, you may also dig:

What is a Voice?–Alexis Deighton MacIntyre

Mr. and Mrs. Talking Machine: The Euphonia, the Phonograph, and the Gendering of Nineteenth Century Mechanical Speech – J. Martin Vest

Only the Sound Itself?: Early Radio, Education, and Archives of "No-Sound"–Amanda Keeler