

Beyond the Every Day: Vocal Potential in AI Mediated Communication

Amina Abbas-Nazari

Blog post published May 22, 2023 in *Sounding Out!*

<https://doi.org/10.59350/hdkc2-hrd04>

Abstract

In summer 2021, sound artist, engineer, musician, and educator Johann Diedrick convened a panel at the intersection of racial bias, listening, and AI technology at Pioneerworks in Brooklyn, NY. Diedrick, 2021 Mozilla Creative Media award recipient and creator of such works as *Dark Matters*, is currently working on identifying the origins of racial bias in voice interface systems. *Dark Matters*, according to

Keywords

Music (Subfield), Article, Artificial Intelligence, Cultural Studies, Digital Media, Gender

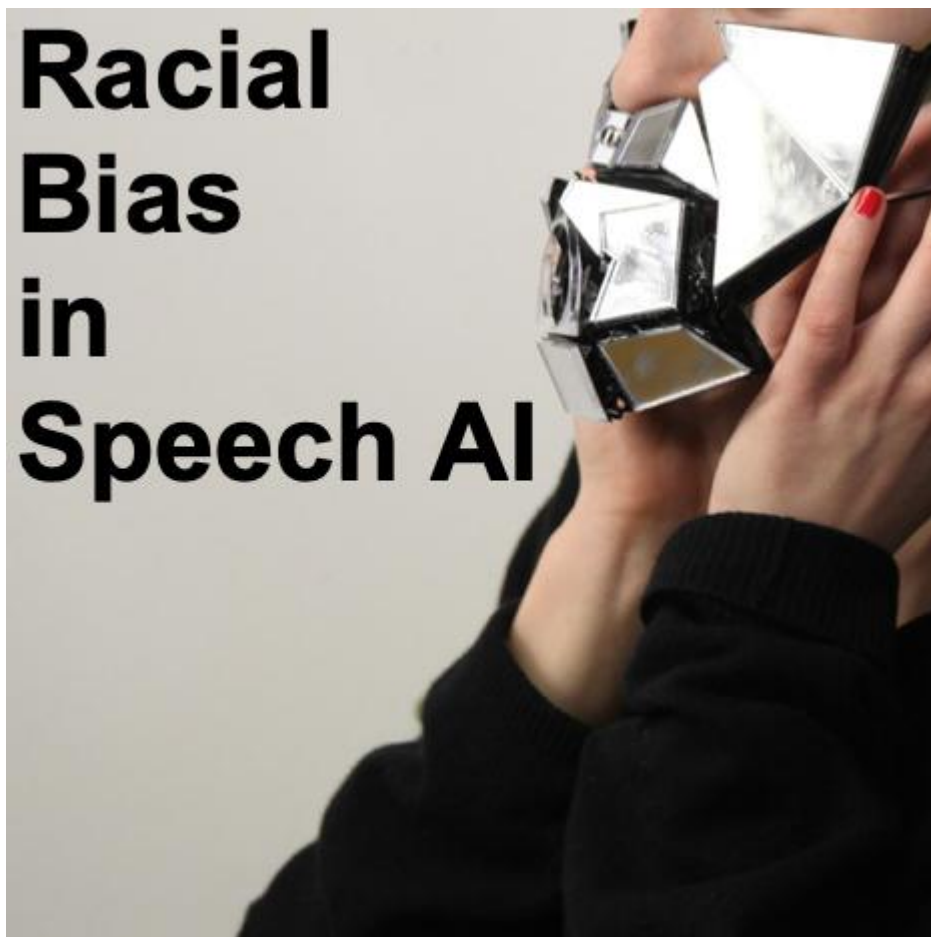


Recommended citation

Abbas-Nazari, A. (2023, May 22). Beyond the Every Day: Vocal Potential in AI Mediated Communication. *Sounding Out!*. <https://doi.org/10.59350/hdkc2-hrd04>

Copyright

Copyright © Amina Abbas-Nazari 2023. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



In summer 2021, sound artist, engineer, musician, and educator Johann Diedrick convened a panel at the intersection of racial bias, listening, and AI technology at Pioneerworks in Brooklyn, NY. Diedrick, 2021 [Mozilla Creative Media](#) award recipient and creator of such works as [Dark Matters](#), is currently working on identifying the origins of racial bias in voice interface systems. *Dark Matters*, according to Squeaky Wheel, "exposes the absence of Black speech in the datasets used to train

voice interface systems in consumer artificial intelligence products such as Alexa and Siri. Utilizing 3D modeling, sound, and storytelling, the project challenges our communities to grapple with racism and inequity through speech and the spoken word, and how AI systems underserve Black communities." And now, he's working with SO! as guest editor for this series for Sounding Out! (along with ed-in-chief JS!). It starts today, with **Amina Abbas-Nazari**, helping us to understand how Speech AI systems operate from a very limiting set of assumptions about the human voice—are we training it, or is it actually training us?

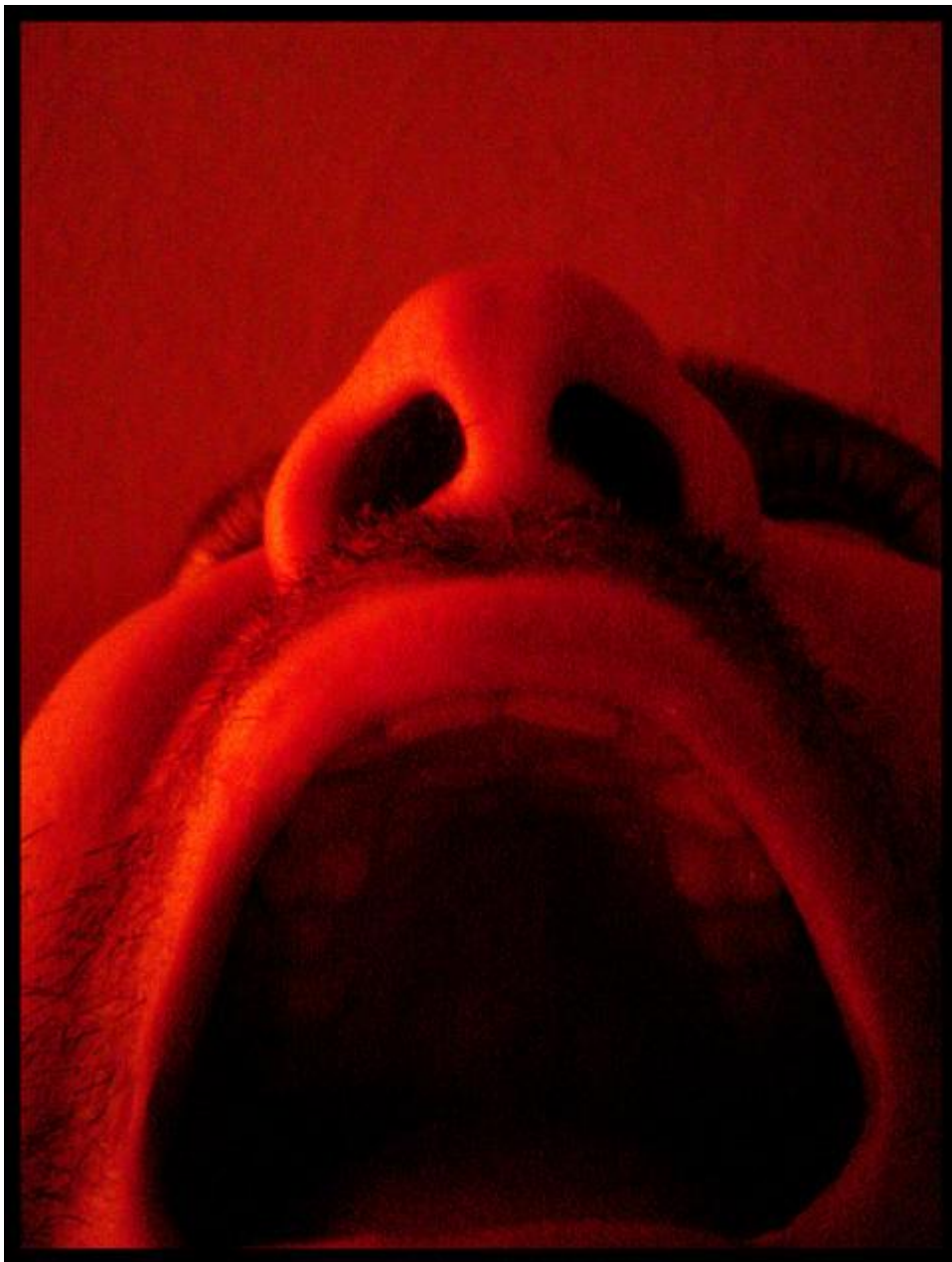
—

Hi, good morning. I'm calling in from Bangalore, India." I'm talking on speakerphone to a man with an obvious Indian accent. He pauses. "Now I have enabled the accent translation," he says. It's the same person, but he sounds completely different: loud and slightly nasal, impossible to distinguish from the accents of my friends in Brooklyn.

The AI startup erasing call center worker accents: is it fighting bias – or perpetuating it? (Wilfred Chan, 24 August 2022)

This telephone interaction was recounted in [The Guardian](#) reporting on a Silicon Valley tech start-up called Sanas. The company provides AI enabled technology for real-time voice modification for call centre workers voices to sound more "Western". The company describes this venture as a solution to improve communication between typically American callers and call centre workers, who might be based in countries such as Philippines and India. Meanwhile, research has found that major companies' AI interactive speech systems exhibit considerable [racial imbalance when trying to recognise Black voices](#) compared to white speakers. As a result, in the hopes of being better heard and understood, Google smart speaker users with regional or ethnic American accents relay that they find themselves [contorting their mouths](#) to imitate Midwestern American accents.

These instances describe racial biases present in voice interactions with AI enabled and mediated communication systems, whereby sounding 'Western' entitles one to more efficient communication, better usability, or increased access to services. This is not a problem specific to AI though. Linguistics researcher [John Baugh](#), writing in 2002, describes how [linguistic profiling](#) is known to have resulted in housing being denied to people of colour in the US via telephone interactions. [Jennifer Stoecker's](#) *The Sonic Color Line* (2016) presents a cultural and political history of the racialized body and how it both informed and was informed by emergent sound technologies. AI mediated communication repeats and reinforces biases that pre-exist the technology itself, but also helping it become even more widely pervasive.



"pain" by Flickr user Pol Neiman ([CC BY-NC-ND 2.0](#))

Mozilla's commendable [Common Voice](#) project aims to 'teach machines how real people speak' by building an open source, multi-language dataset of voices to improve usability for non-Western speaking or sounding voices. But singer and musicologist, [Nina Sun Eidsheim](#) describes how 'a specific voice's sonic potentiality [in] its execution can exceed imagination' (7), and voices as having 'an infinity of unrealised manifestations' (8) in [The Race of Sound](#) (2019). Eidsheim's sentiments describe a vocal potential, through musicality, that exists beyond ideas of accents and dialects, and vocal markers of categorised identity. As a practicing vocal performer, I recognise and resonate with Eidsheim's ideas I have a particular interest in extended and experimental vocality, especially gained through my time singing with [Musarc](#) Choir and working with artist [Fani Parali](#). In these instances, I have experienced the pleasurable challenge of being asked to vocalise the mythical, animal, imagined, alien and otherworldly edges of the sonic sphere, to explore complex relations between bodies, ecologies, space and time, illuminated through vocal expression.



Joy by Flickr user François Karm, cropped by SO! (CC BY-NC 2.0)

Following from Eidsheim, and through my own vocal practice, I believe AI's prerequisite of voices as "fixed, extractable, and measurable 'sound object[s]' located within the body" is over-simplistic and reductive. Voices, within systems of AI, are made to seem only as computable delineations of person, personality and identity, constrained to standardised stereotypes. By highlighting vocal potential, I offer a unique critique of the way voices are currently comprehended in AI recognition systems. When we appreciate the voice beyond the homogenous, we give it authority and autonomy, ultimately leading to a fuller understanding of the voice and its sounding capabilities.

My [current PhD research](#), *Speculative Voicing*, applies thinking about the voice from a musical perspective to the sound and sounding of voices in artificially intelligent conversational systems. Herby the voice becomes an instrument of the body to explore its sonic materiality, vocal potential and extremities of expression, rather than being comprehended in conjunction to vocal markers of identity aligning to categories of race, gender, age, etc. In turn, this opens space for the voice to be understood as a shapeshifting, morphing and malleable entity, with immense sounding potential beyond what might be considered ordinary or everyday speech. Over the long term this provides discussion of how experimenting with vocal potential may illuminate more diverse perspectives about our sense of self and being in relation to vocal sounding.

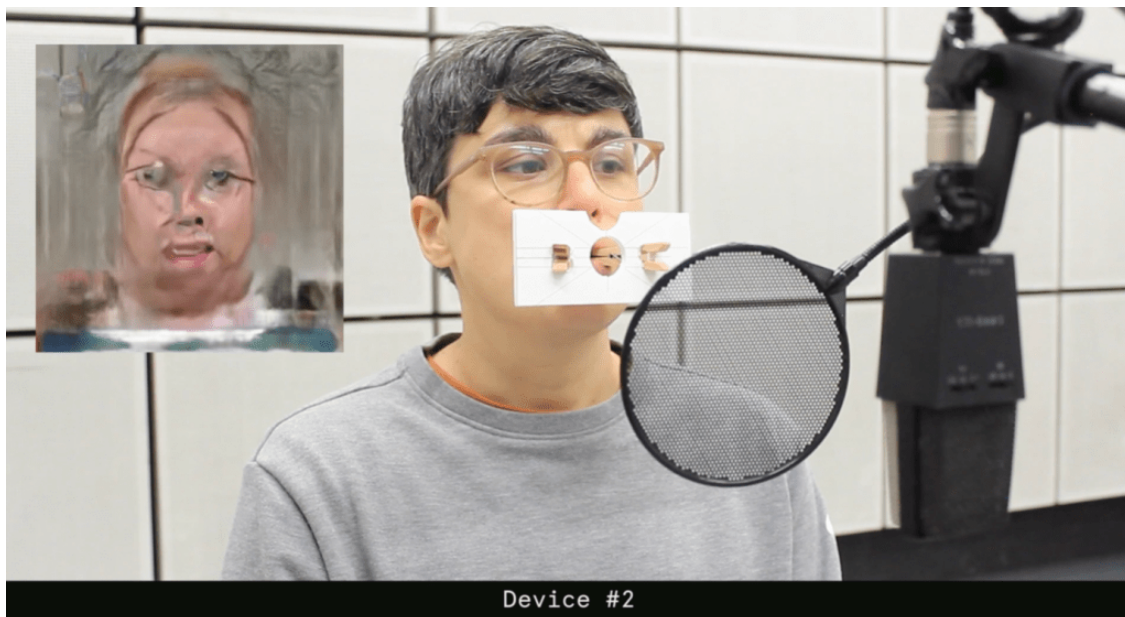
Vocal and movement artist [Elaine Mitchener](#) exhibits the disillusion of the voice as 'fixed' perfectly in her performance of Christian Marclay's *No!*, which I attended one hot summer's evening at the London Contemporary Music Festival in 2022. [Marclay's graphic score](#) uses cut outs from comic book strips to direct the performer to vocalise a myriad of 'No's.



In connection with Fraenkel Gallery's 2021 exhibition, experimental vocalist Elaine Mitchener performs Christian Marclay's graphic score, "No!" Image by author.

Mitchener's rendering of the piece involved the cooperation and coordination of her entire body, carefully crafting lips, teeth, tongue, muscles and ligaments to construct each iteration of 'No.' Each transmutation of Mitchener's 'No's' came with a distinct meaning, context, and significance, contained within the vocalisation of this one simple syllable. Every utterance explored a new vocal potential, enabled by her body alone. In the context of AI mediated communication, we can see this way of working with the voice renders the idea of the voice as 'fixed' as redundant. Mitchener's vocal potential demonstrates that voices can and do exist beyond AI's prescribed comprehension of vocal sounding.

In order to further understand how AI transcribes understandings of voice onto notions of identity, and vocal potential, I produced the practice project *Polyphonic Embodiment(s)* as part of my PhD research, in collaboration with Nestor Pestana, with AI development by Sitraka Rakotoniaina. The AI we created for this project is based upon a speech-to-face recognition AI that aims to be able to tell what your face looks like from the sound of your voice. The prospective impact of this AI is deeply unsettling, as its intended applications are wide-ranging – from entertainment to security, and as previously described AI recognition systems are inherently biased.



Still from project video for Polyphonic Embodiment(s). Image by author.

This multi-modal form of comprehending voice is also a hot topic of research being conducted by major research institutions including Oxford University and Massachusetts Institute of Technology. We wanted to explore this AI recognition programme in conjunction with an understanding of vocal potential and the voice as a sonic material shaped by the body. As the project title suggests, the work invites people to consider the multi-dimensional nature of voice and vocal identity from an embodied standpoint. Additionally, it calls for contemplation of the relationships between voice and identity, and individuals having multiple or evolving versions of identity. The collaboration with the custom-made AI software creates a feedback loop to reflect on how peoples' vocal sounding is "seen" by AI, to contest the way voices are currently heard, comprehended and utilised by AI, and indeed the AI industry.

<https://vimeo.com/784768626>

The video documentation for this project shows 'facial' images produced by the voice-to-face recognition AI, when activated by my voice, modified with simple DIY voice devices. Each new voice variation, created by each device, produces a different outputted face image. Some images perhaps resemble my face? (e.g. Device #8) some might be considered more masculine? (e.g. Device #10) and some are just disconcerting (e.g. Device #4). The speculative nature of *Polyphonic Embodiment(s)* is not to suggest that people should modify their voices in interaction with AI communication systems. Rather the simple devices work with bodily architecture and exaggerate its materiality, considering it as a flexible instrument to explore vocal potential. In turn this sheds light on the normative assumptions contained within AI's readings of voice and its relationships to facial image and identity construction.

Through this artistic, practice-led research I hope to evolve and augment discussion around how the sounding of voices is comprehended by different disciplines of research. Taking a standpoint from music and design practice, I believe this can contest ways of working in the realms of AI mediated communication and shape the ways we understand notions of (vocal) identity: as complex, fluid, malleable, and ultimately not reducible to Western logics of sounding.

Featured Image: Still image from Polyphonic Embodiments, courtesy of author.

Amina Abbas-Nazari is a practicing speculative designer, researcher, and vocal performer. Amina has researched the voice in conjunction with emerging technology, through practice, since 2008 and is now completing a PhD in the School of Communication at the Royal College of Art, focusing on the sound and sounding of voices in artificially intelligent conversational systems. She has presented her work at the London Design Festival, Design Museum, Barbican Centre, V&A, Milan Furniture Fair, Venice Architecture Biennial, Critical Media Lab, Switzerland, Litost Gallery, Prague and Harvard University, America. She has performed internationally with choirs and regularly collaborates with artists as an experimental vocalist



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

What is a Voice?—Alexis Deighton MacIntyre

Voice as Ecology: Voice Donation, Materiality, Identity—Steph Ceraso

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

One Scream is All it Takes: Voice Activated Personal Safety, Audio Surveillance, and Gender Violence—María Edurne Zuazu

Echo and the Chorus of Female Machines—AO Roberts

On Sound and Pleasure: Meditations on the Human Voice— Yvon Bonefant