

Labor, Power, and Belonging: The Work of Voice in the Age of AI Reproduction

Shm Almeda
University of California, Berkeley
Berkeley, CA, USA
shm.almeda@berkeley.edu

Robin Netzorg
University of California, Berkeley
Berkeley, CA, USA
robert_netzorg@berkeley.edu

Isabel Li
University of California, Berkeley
Berkeley, CA, USA
isabel.yueli@gmail.com

Ethan Tam
University of California, Berkeley
Berkeley, CA, USA
etam1@berkeley.edu

Skyla Ma
University of California, Berkeley
Berkeley, CA, USA
skylama1029@berkeley.edu

Bob Tianqi Wei
University of California, Berkeley
Berkeley, CA, USA
bobwei3598@gmail.com

Abstract

Developments in Speech AI have facilitated a rise in text-to-speech (TTS) generation and voice-cloning systems that progressively require less data to synthesize higher-quality speech. As voice actors adapt to the impacts of voice-cloning and speech data collection on their industry, the researchers and developers driving these changes are removed from this affected context. Towards an informed perspective that centers a community directly impacted by Speech AI development and deployment, we interview 15 professional voice artists on their perceptions and responses to the emergence of Speech AI in their work. We identify how access to community, prior experiences negotiating creative labor rights, and existing exploitative power structures influence artists' expectations and agency in the face of AI-driven disruption. We call for technologists to take a community-led approach, identifying opportunities to shift Speech AI towards creative empowerment, and to look beyond technology-centered solutions towards shaping a more equitable future for voice work.

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**; • **Computing methodologies** → *Artificial intelligence*.

Keywords

Voice acting, generative AI, speech generation, voice-cloning, Speech AI, creative precarity, AI impact on creative communities, creative labor

ACM Reference Format:

Shm Almeda, Robin Netzorg, Isabel Li, Ethan Tam, Skyla Ma, and Bob Tianqi Wei. 2025. Labor, Power, and Belonging: The Work of Voice in the Age of AI Reproduction. In *The 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)*, June 23–26, 2025, Athens, Greece. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3715275.3732082>



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

FAccT '25, Athens, Greece

© 2025 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-1482-5/25/06

<https://doi.org/10.1145/3715275.3732082>

1 Introduction

Advances in machine learning have enabled the rising popularity of text-to-speech (TTS) applications, which allow users to generate audio of voices speaking an inputted phrase. These advances have also led to *voice-cloning* systems, capable of mimicking a *specific* person's voice, given a sample of them speaking. With each advancement, these systems require less data (e.g. only a few seconds of the person speaking) and can generate progressively higher-quality imitations.

Misuse of AI systems has become a major point of contention in disputes between creative labor unions and the entertainment industry, including concerns that generative AI will be used to replace jobs and non-consensually clone actors [43–45]. Meanwhile, AI technology researchers and developers are removed from this affected context, even as their work plays a significant, and potentially harmful, role in this space [8, 24].

In this work, we center and amplify the perspectives of a community actively navigating and negotiating the impacts of Speech AI technology on their work: the voice acting and performance community.

Toward an informed perspective that centers directly impacted voices, we interviewed 15 professional voice artists actively working, auditioning, and performing in various industries. We reporting on their experiences, perceptions and responses to the emergence of Speech AI in their work. We focus on the impacts of voice data collection practices and of voice reproduction applications, to consider how both the accelerating development and deployment of Speech AI systems is changing the shape of voice work.

We characterize the entangled nature of contemporary voice work, where existing power imbalances and patterns of exploitation are being exacerbated and disrupted by these emergent technologies. We report on the nature of voice performance as personal and embodied, and how navigating the circumstances of their labor characterizes the daily reality of voice workers, who must routinely negotiate their individual and collective rights as performers. We find that Speech-AI development is driving practices that undermine needs that voice artist communities and labor organizations are actively advocating for.

We use these findings to ground a discussion on the role that AI developers, researchers, and policymakers can play in challenging, rather than perpetuating, the ongoing exploitation and precarity of creative labor. We identify opportunities to design for creative

empowerment by centering voice artists in the development of *power-conscious Speech AI*. Beyond calling for the direct involvement of voice workers and their community organizations in the development of new technologies, we also identify the need to contribute to efforts beyond techno-solutionism, and to honor the expertise of voice artists as creative labor rights negotiators. We call on developers, researchers, and policymakers to act in solidarity with the creative community’s existing organizational efforts to understand, articulate and defend the needs of voice artists, and other workers in creatively precarious positions, towards the design of a more equitable future for voice work.

2 Background and Related Work

2.1 The Unique Landscape of Speech AI

Machine learning and generative text and video models have seen an explosive increase in interest over the past few years [21, 38]. While State-of-the-Art (SOTA) image, video, and text generation systems struggle with problems of compositionality [33], hallucinations [42], and generalization [52], the problem of automatically generating arbitrary speech from any individual’s voice is essentially solved for the world’s most common languages. Prior work has shown that untrained listeners cannot reliably distinguish between audio produced by SOTA voice cloning models, and the original voices they are emulating [18, 36], and that these models outputs are only readily distinguishable to voice experts and with automatic methods [9, 34, 46]. Moreover, SOTA large deep learning-based TTS systems can generate an accurate clone with less than 10 seconds of audio [19, 27, 53, 55].

The recent success in TTS systems is due in part to speech’s highly structural nature resulting in sparse and efficient codecs [15, 27]. Since the first TTS system’s release in 1968, speech processing researchers and practitioners have been able to generate intelligible speech from an array of different speakers [29]. While many open problems remain, such as controllable TTS [54], real-time generation [35], and generating more diverse, expressive, and conversational-like speech [28, 37, 57], current methods are more than capable of generating realistic voices for personal, commercial, and malicious ends.

2.2 Harms of Speech AI

Much work has been done on assessing and ameliorating harms caused by automatic systems [13], and how they fail to center the needs of impacted human communities [51]. Despite much focus on synthetic speech detection in the speech processing community, Speech AI’s unique location in the modern landscape has led to a body of work evaluating the potential for broader harms and ethical concerns from its nascency. While much prior work has explored the harms and biases resulting from clearly synthetic voices and digital assistants [30, 48], less than 10% of papers on generative speech systems from 2018-2023 discussed potential negative impact [8]. Likewise, in an audit of audio datasets, Agnew et al. found that 20% of the data contained in 80 speech datasets is explicitly copyright-infringing, with more being used for speech synthesis without the speaker’s knowledge or consent [3].

Hutiri et al. contributed a taxonomy of speech AI harms for members of the general public at large, including voice-cloning incidents

impacting voice actors [24]. However, prior work has largely focused on the misuse of deployed Speech AI application towards privacy and impersonation concerns, rather than investigating the impact of these systems on the voice artists they collect speech data from, and the labor impact of voice-reproduction systems displacing voice artists’ work opportunities. While voice artists may be subject to the same data misuse risks as other members of the general public, proposed privacy measures may not apply to them, as individuals whose work often entails widely and publicly sharing recordings of their voice.

We seek to build off this body of work to contribute a better understanding of how speech AI’s harms are impacting the voice artist community, and the unique considerations and protections this community may need in response.

2.3 On the Impacts of AI and Automation on Labor

Recent work has called for researchers and developers of data-driven technologies to shift away from a view that positions data as apolitical and objective, towards recognizing the implications of datasets as collections of political artifacts extracted from human contexts, often exploitatively [31, 47]. Applying this human-centered understanding of data collection frames the development of contemporary AI image generation models as extractive systems that profit from visual artists’ labor, without due consent or compensation [7, 22, 26].

Prior work has also investigated the impact that rapid deployment of automation technologies can have on workers forced to adapt to these new systems in real-world situations, such as in public transportation [5], in essential custodial and waste work [20], and in creative industries [14, 22, 26]. These studies often showcase the ways that technologies developed outside of the real-world-contexts where they will be deployed tend to operate on a myopic, reductionist understanding of very work they are trying to automate—ultimately disrupting and obscuring workers livelihoods rather than supporting them. AI-driven disruptions often exacerbate existing inequalities, while forcing affected workers to rapidly adapt to the world created by AI and its developers, all without their appeal, consent, or acknowledgment [6].

We seek to extend and apply these lines of study to the voice artist community, and to consider the emergent impacts of AI technology on the shape of their work. We take interest in the multifaceted role that voice artists play: as creative practitioners navigating already exploitative industries, and as workers whose livelihoods stand to be directly disrupted — by the *development* of Speech AI driving the extractive reuse of performers’ voice data, and by the *deployment* of voice-reproduction technologies driving labor displacement in their industries.

3 Method

To achieve a perspective on Speech AI development that centers directly impacted voices, we interviewed 15 voice actors (VAs) actively working in the voice industry on their perceptions and responses to Speech AI.

The interviews were conducted in a semi-structured manner, with one to four authors present for each interview. Interviews

lasted 60 minutes and were conducted over Zoom through April to September 2024.

3.1 Protocol

We designed a semi-structured interview protocol to evaluate our participants' perceptions and responses to emerging Speech AI technologies and development practices, especially with regard to TTS, voice cloning, and speech data collection. The protocol consisted of three sections: introductory questions on their prior experiences, demonstrations of emergent technologies, and a "traffic light exercise." Our full protocol is available in Appendix A.

3.1.1 Demonstrations of Existing Technologies. As many of our participants lacked direct experience with Speech AI technology, their initial perceptions could be heavily based in speculation or secondhand accounts. To counteract this and ground discussions of the technology in a shared understanding, we used screen-sharing to present concrete demonstrations of contemporary speech AI systems: voice cloning TTS [56], text prompt-based voice modification and synthesis [53], and Speech AI-assisted editing tools. Participants were invited to ask questions, and then to discuss their impressions of the system and its capabilities.

3.1.2 Traffic Light Exercise. The session concluded with an exercise designed to capture participants' preferences, expectations, and potential negotiation practices regarding the introduction of AI into their work.

We prepared six scenarios that were each grounded in ways that Speech AI-driven trends might impact their future work. For example: "Scenario 2: A company wants to pay you to acquire your voice data to use in a large training dataset." We read each of these six scenarios aloud while sharing a graphic with the participant that described the four "colors" they could respond with, where each color indicates a level of consent (See Fig.2). Red is a hard "no" under any circumstances; Green is "enthusiastic consent"; we offered Yellow and Blue as "with major limitations" and "with minor limitations" respectively to help participants with nuanced perspectives better differentiate their response, and to encourage critical discussion on the kinds of boundaries and preferences they might negotiate, and the considerations involved in their decision. In designing this exercise, we drew directly from Strengers et al. in *What can HCI learn from sexual consent?* [49], especially the "[T]raffic light system" component of the TEASE guideline.

3.2 Recruitment

We recruited participants via social media and public voice-cast information sites like IMDB¹. We utilized snowball recruiting, asking enrolled subjects to pass information about the study to their community. This strategy significantly improved response rates; voice community members are reasonably wary of connecting with AI researchers. After 3 interviews, we met with members of the Alliance of Canadian Cinema, Television and Radio Artists (ACTRA), who then shared our screener survey in their community meetings and mailing list. Gaining trust within the community increased participants' willingness to respond and schedule interviews. We discuss the potential limitations of this recruitment strategy in 3.4.

¹<https://www.imdb.com/>

We interviewed 15 voice artists in total. All participants were either from Canada or the United States, with 7 identifying as men, 6 as women, and 1 non-binary individual. Full participant details are recorded in Table 1. Participants were compensated with a gift card valued at 20 USD. Our Institutional Review Board reviewed and approved our protocol.

3.3 Analysis

Our analysis followed a qualitative reflexive thematic analysis approach [11, 12]. Authors reviewed all interview recordings and manually corrected transcripts for accuracy. Each interview was analyzed by at least two authors. We conducted an open coding process [17]; codes were iteratively refined and clustered into thematic categories, using post-it notes in an online collaborative whiteboard platform.

We recorded structured data from our traffic light exercise, which provided participants' responses to potential AI use case scenarios (See Fig. 1). All codes and preliminary themes were collaboratively reviewed and discussed by the research team across three synchronous meetings to ensure consistency and interpretive depth. Through consensus, we synthesized these themes into a final set of overarching narratives that underpin our findings.

3.4 Limitations

Our sample size of 15 voice workers is far from a total representation of the voice work community. In particular, while snowball recruiting allowed us to improve response rates and establish credibility and rapport with voice workers reasonably wary of community outsiders, this resulted in a participant pool skewed towards a particular community, centered around Canadian union members. As a result, our study should not be considered representative of the reality for the global community of voice actors; we hope that future work could bring forth the voices and unique challenges faced by more non-union actors and by voice workers in other regions and industry contexts.

4 Findings

4.1 The work of voice

In this section, we report on the present nature of voice labor, to contextualize our understanding of this space that Speech AI is actively disrupting.

4.1.1 The work of voice performance. Despite describing our participants as voice artists, our participants' work is sociotechnically entangled across industries and tasks. All participants except P1 had experience working in multiple roles, and for different sub-industries. Experiences spanned across a variety of industries, including film, video games, audiobooks, and dramatic podcasts. Experiences were not limited to entertainment industries; beyond commercial advertising work, participants also noted working or auditioning for roles like, narrating corporate training videos, recording for automated phone answering systems, creating accessible audio descriptions, and – more recently – providing recordings of their voice as training data for a speech application. Many of our participants, especially the 7 who identified as experts (P3, P4, P5, P6, P8, P9, P13) have worked with clients in more distinct roles than they

P#	Age Range	Experience (Years)	Voice Industry Role(s) / Experience	Skill Level	Union Involvement
P1	18-24	1-2	Voice actor (VA).	Moderate Skill	None reported
P2	18-24	1-2	VA, podcast production, student	Moderate Skill	None reported
P3	30-39	3-10	Actor & VA for film, TV, animation	Expert	Active
P4	50-59	20+	VA, voice studio founder, organizer	Expert	Active, Organizer
P5	60+	20+	Retired commercial voice actor; writer	Expert	Active
P6	30-39	10-20	VA for commercials, animation, games	Expert	Active, Organizer
P7	30-39	<1	Theatre actor, VA for film	Novice	Limited
P8	30-39	10-20	VA for animation and games, theater	Expert	Active
P9	50-59	20+	Theatrical VA	Expert	Active
P10	40-49	3-10	Artist, actor, screenwriter, filmmaker	Moderate skill	Limited
P11	30-39	3-10	Commercial voiceover, animation VA	Very skilled	Active, Organizer
P12	50-59	20+	Theatre, singing, audio description work	Very skilled	Active
P13	60+	20+	Union organizer, actor, dialect coach	Expert	Active, Organizer
P14	30-39	10-20	VA for animation, director	Very skilled	Active
P15	40-49	<1	Actor for TV and film, emerging VA	Novice	Limited

Table 1: Data as self-reported by our 15 participants. There is a wide range of positions and tasks across industries that employ voice work. Most of the expert VAs we interviewed held more distinct roles than we are able to list here.

could feasibly list. See Table 1 for a complete overview of our 15 participants.

Despite the public and commercialized nature of the industries they work in, many participants described experience the work of voice and voice artistry as personal, creative, and embodied. Voice is a conduit for their individual identity, creativity, style, and story.

“One of the whole points of doing voice over performance and like, all the studying and the training that we do is like to be able to manipulate your voice in different ways...it’s a passion, right? Like it’s not, ‘you just need to get the job done.’ ...There’s a different kind of care to it. There is something kind of magical about it.” -P11

Participants mentioned a general monetary devaluation of performance, resulting in a “Passion Tax” whereby performers are expected to do certain jobs for low rates or possibly free (P6), echoing findings from prior work on creative precarity [14].

While we focused on recruiting voice *performers*, some of our participants also represented crucial voice “support” roles, like studio engineer (P4) and vocal coach (P13). Participants noted an increased expectation for actors to complete support role tasks, with 10/15 describing recording and mixing their own audio at home. Since the Covid 19 pandemic, the voice industry has transitioned away from in-person voice-booth performances and towards “self-taping” for auditions and gigs (P11). While participants enjoyed the career flexibility afforded by increasingly accessible end-user audio technologies, and were enthusiastic that lower barriers to entry enable independent creatives to produce more projects with less resources (e.g. P2’s theatrical podcast), participants stressed the unique value of collaborating with other creatives in person:

“I enjoy being in the voice booth... I like working with the writer and working with the brand and the technician. We’re all creating something together.” -P6

Te accessibility of editing technologies is driving expectations for performers to work remotely and complete audio technicians tasks themselves. This trend threatens to eliminate opportunities to collaborate with others in person; this isolates artists, making it harder for them to build supportive and creative relationships with others. P12 lamented that what clients gain from shifting labor onto actors rarely translates into increased compensation:

“I’m doing triple the work. But I’m still getting paid the same... and they save money because they don’t have to pay for the engineer. [...] Technology is getting better and faster. So in theory that should free up resources....where is that extra money going? It’s certainly not going to the actors.” - P12

As we asked participants about the ways that new technologies, like Speech AI, are surfacing in their spaces of work, conversations quickly became about topics like this — the socioeconomic dynamics of voice labor that VAs must regularly navigate, far beyond the tasks of voice performance itself, which we address next.

4.1.2 The work of navigating and negotiating voice labor. Precarity and power imbalance are mainstays of the entertainment and service industries where voice artists find employment, leading them to develop expertise in routinely negotiating the terms of their work.

Figure 1 shows results from the “traffic light exercise” that concluded each interview, arranged by participant. Figure 2 displays the same data, but arranged to comparison responses across the six scenarios. Across our findings, we report on participants’ responses

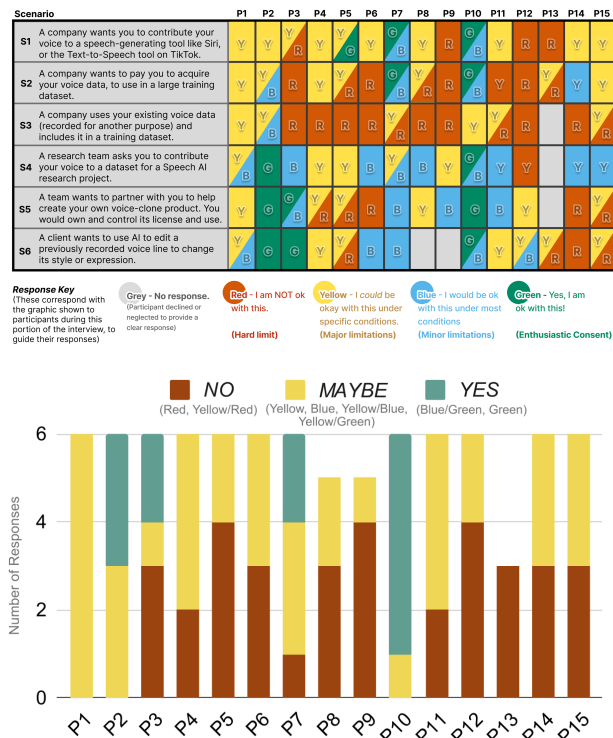


Figure 1: Traffic light exercise results, organized per participant. Two colors indicate where a participant’s response was “between” two options (e.g. “I could go yellow or blue on this.”) For the bar chart, we summarize results into 3 groups, “No” (Red and Yellow/Red), “Maybe” (Yellow, Blue, Yellow/Blue, or Yellow/Green) and “Yes” (Blue/Green, Green). (Based on the conversation, we chose to count P5’s “Yellow/Green” response as a “Maybe”.)

to these scenarios and the limitations, considerations, and boundaries they surfaced, contextualized with data from throughout each interview.

When discussing how they would navigate potential future scenarios, participants frequently referenced past experiences negotiating their rights and terms as performers. These anecdotes helped illustrate how existing labor dynamics impact their expectations for the future of their work, and how having the skills to adapt to these changing dynamics characterizes the present reality of voice work. VAs must routinely complete tasks beyond voice performance like: finding and securing fair working opportunities, preparing and submitting auditioning materials, navigating relationships with agents, clients, corporations, and licensees, negotiating the terms of their performance contracts or licenses, and working to enact and defend their rights and preferences before, during, and sometimes long after a particular performance,

“We license the sound of our voice to a company for a product in a certain medium for a certain amount of time.” -P4

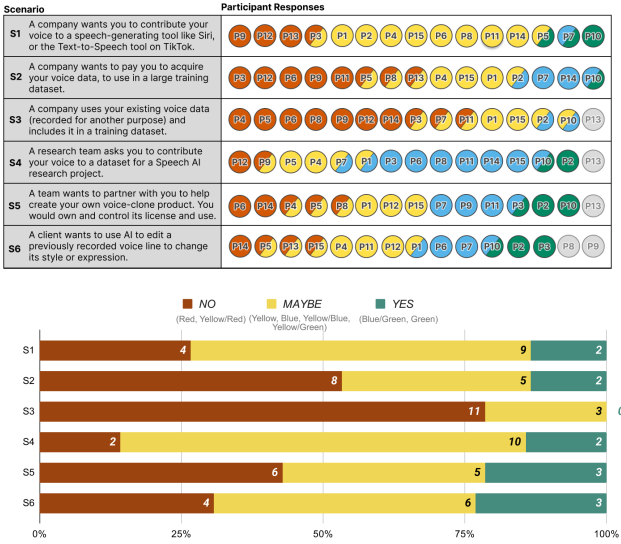


Figure 2: Traffic light exercise results, organized by scenario. We use the same summarization categories as described in the caption for Figure 1.

Compensation was the most frequently cited consideration that participants discussed in their responses to various potential scenarios, mentioned as critical by 10/15 participants. 4/15 discussed how *termed contracts* are a desirable model for securing compensation, consent, and control: these agreements allow VAs to agree to a specified use of their recorded voice performance for a limited amount of time, for pay to be distributed throughout that time period and/or upon contract renewal, and for there to be an option to withdraw participation at the end of each term.

“When I booked, like, a McDonald’s commercial I do the voice for, they say, ‘We’re going to use this for one year.’ When I do that, that means I’m not allowed to then get a job voicing for Burger King, Wendy’s, any other competitor.” - P11

6/15 participants mentioned *informed usage control* as important to them; they want to be fully informed on how their recorded voice data, and any system they might contribute to, will be used, and to have opportunities to withdraw if the usage changes.

5/15 participants discussed the role of “trust” — or *distrust*, especially in the case of technology companies — when negotiating and considering the safety of potential work opportunities. Participants referenced widespread mistrust of tech companies, citing their reputation for nonconsensual data use and for prioritizing profit over community interests.

“I’m always hesitant to trust any company. Personally, I think every company will always focus on making money. So there’s always just a little bit of like a gray area [...] Artists and scientists kind of come at odds with each other because of pushing forward technology versus humanity and empathy.”

4.1.3 The work of voice community organization & outreach. Our 15 participants represented across at least 3 voice labor organizations, including ACTRA², SAG-AFTRA³, and NAVA⁴ (a non-union voice acting association). 10/15 (66%) of participants were actively involved in a labor group, including 4 (P4, P6, P11, P13) with organizational roles. Of the remaining five, three (P7, P10, P15) described themselves as members with limited involvement.

Participants did note that typically, professional voice acting experience and example materials are required to join a labor union, creating barriers for novices to access these benefits.

Union agencies negotiate higher pay rates for members and provide benefits like health insurance, a pension fund, and access to a community network.

“Union, you get paid on the scale [...] without the union, you have no rights and no protection.” -P10

“Before I joined the union, I did a lot of non-union work... the difference is palpable. I mean, a level of disrespect and disregard ... People are not taken care of, people are not fed properly.” - P11

VA work is frequently transient and independent; performers benefit from the support, information, and solidarity they receive from participating in community.

“It is very isolated... I would be recording at home, so unless I talk to somebody else about, ‘Hey, I’m having a problem with this. Can somebody help me out?’ I would be completely on my own.” - P12

If you’re not union, then you’re just, like, looking through Facebook groups for gigs.” - P10

As union members themselves (especially the four actively serving in organizational roles) participants reported on the many service tasks they take on to support and protect the actor and performer community. These include: attending labor demonstrations and community meetings; crafting informational resources; managing email newsletters, websites, and online seminars to keep the disparate network of union performers informed and involved; helping to prepare, and defend, collective agreements and bargaining terms; meeting with company representatives and lawmakers to represent the community, and more. In subsequent findings sections, we report on how community involvement surface in the ways participants respond to potential AI-driven work disruptions.

4.2 The impacts of Speech AI on voice work

In this section, we report our findings on the various impacts of emergent Speech AI technology development and deployment trends on the dynamics of voice work — how it shaping voice performance, and the dimensions of labor negotiation and organization that characterize the reality of voice work beyond performance.

4.2.1 Impacting the craft and processes of voice performance. Some participants described concern that the adoption of Speech AI technologies is threatening the opportunities they have to engage in their craft as performing artists.

Voice is an intimate, embodied attribute that is personal to every human, and the same can be said of artistic expression, performance, and style. Engaging with the craft of voice performance is what makes this work meaningful and enjoyable for many voice artists. Neglecting to respect the depth of humanness at the core of voice work can lead to the development of new technologies and data practices that are reductionist and dehumanizing.

“Producers are looking at our work simply as data and not as acting or performance... it’s just so condescending.” - P9

VAs are protective of their voices, viewing them as extensions of their own identities, and representations of their artistry.

“We don’t have much as a person to call our own ... my voice is my voice. If you don’t want it, no problem. But don’t have it used elsewhere.” - P5

6/15 participants expressed fear that their voices could be misused for harmful speech generation or to misrepresent them, e.g. for “a political party that I don’t approve of, or [...] something that was violent or abusive” (P12) or for “hacking and scamming” (P5).

“I don’t want my voice being used to say things that I would never say... I don’t think I would be okay with that.” - P6

P8 distinguished between human voice impressions and voice cloning, noting that the latter felt like a “murky” theft of identity:

“When it started to bleed into, oh, they’re taking my voice and manipulating my performance to create new content that I have no say in [...] It feels different than if someone does an impression of you right? Because then there’s still that separation right there, it’s someone pretending to be you, but is clearly *not* you, and cannot represent your thoughts and cannot endorse things on your behalf.” - P8

Participants associated misuse of their voices for harmful purposes with “[taking] away” (P12) their voices and control. P10 stated that it was important to them that their voice is used to represent characters of their ethnicity and/or heritage, and that voice-cloning technologies could enable uses that would erase that identity.

A majority of our participants (10/15) expressed belief that **certain AI tools could make a positive impact** on their work and in their lives. They were particularly enthusiastic to share ideas for AI tools that would center their support needs – for example, by supporting tedious and undesirable elements of their labor like editing, setting up a suitable recording environment, and helping them with rehearsal.

Participants (P1, P10, P12, P13, P15) envisioned AI tools automating the tedious, time-consuming editing processes that actors are now regularly expected to perform. P1 said, “If there was an AI that could remove the clicks that happen in my mouth from all my auditions, that would be so helpful! [...] That would help auditioners out, that would help voice engineers out.”

P12 proposed that AI could enable them to achieve new effects:

“I could use it to change my voice, to be more of a masculine sound, or to [...] speak other languages for you. So that aspect of it is a cool idea that I could [...]”

²<https://actratoronto.com/>

³<https://www.sagaftra.org/>

⁴<https://navavoices.org/>

multiply the output of my own voice to do things, as long as its within my control.” - P12

VAs were more open to contributing to Speech AI research (S4) compared to other potential future scenarios, and to contributing to applications that enable new tasks that human labor previously would not have been used for — as long as their negotiated boundaries were met.

“...narrating the New York Times every day when it comes out, so you have an accessible version. No human could narrate the New York Times cover to cover at the moment that it’s released[...] if New York Times wanted to license my voice, and I knew explicitly that my voice is going to be used for that, and had a licensing deal, then that’s great! You could synthesize my voice.” - P4

We showed participants AI voice editing applications (like Descript⁵) that advertise the use of voice cloning to seamlessly make small edits to otherwise human-recorded audio, e.g. to remove filler words. Many participants were positive about using such tools for minor edits. P2 was enthusiastic about these tools as a director themselves, but stated that they would lose respect for a director who used AI excessively to change many lines of dialogue, as this would represent a failure in creative direction.

Participants discussed how easily these tools could replace the function of “pickup” sessions, where a voice actor re-records a line to fix mistakes or change its style; this may require them to return to a physical studio.

“[If] I’m away when the message comes in for those pickups, and they want it done right away, and I’m not near my computer. I could use AI to do that.” - P12

P4 was open to this kind of AI use, as long as they would still receive some form of pickup fee. Meanwhile, P8 expressed optimism that AI editing technologies could reduce the need for pickup session fees, lowering barriers for smaller companies to hire voice actors more often.

That said, 4/15 participants explicitly distinguished between feeling comfortable with AI being used for practical edits, and uncomfortable with it replacing artistic tasks, like significantly altering the style of a performance. P13 worried that new technology “threatens to undermine the artistry by prioritizing efficiency over authenticity.”

“I think AI should be used to help in the creation of art. but should not be used to create art.” -P9

Participants (P9, P11, P14, P15) explicitly stated that they would prefer to return to the studio for a pick-up session, *even if they were offered the same compensation, for no additional work, if they agreed to have the line edited with AI instead.*

“I want to go in ... do pickups. I love doing that. While I think it’s important absolutely that the 3 C’s⁶ are adhered to, I still don’t want that work to be done for me digitally.” - P9

⁵<https://www.descript.com/>

⁶P9 is referring to the Consent, Control, and Compensation components of the fAIR voices framework as distributed by NAVA[40], as discussed in section 4.2.3.

4.2.2 Disrupting the landscape and power dynamics of voice work. Speech AI is emerging as a disruptive influence in spaces that already frequently put voice actors into precarious or exploitative positions. For instance, 6/15 participants cited AI as a factor impacting declining work opportunities. P10 noted that they “haven’t had a voice acting audition in five months... I used to get at least one a week [...] AI is probably taking over voice acting budgets...” P12 had been recording accessible audio descriptions for TV and film for 20 years when, in 2020, her client shut down — selling its contracts to a new company with no intention of using human labor:

“I contacted this new audio description company, they said, ‘Oh, yeah, we do text to speech. So somebody writes the script and then it’s an AI [...] that does it.’ So that meant that that was a door that was now closed to me.” - P12

A lack of transparency around AI use makes it difficult for voice workers to evaluate the future of their work; the subsequent instability and uncertainty is disempowering, undermining their ability to effectively prepare or respond to what may be a significant disruption to their ongoing career.

“...the bigger damage that we’re seeing is that we’re not necessarily knowing if jobs have been taken by AI, because those jobs are just no longer being put out to the market.” - P4

Participants (P3, P7, P10) suggested how existing power dynamics with clients or agents could exert pressure on their choices around AI use.

When asked for their response to S6 (*A client wants to use AI to edit a previously recorded voice line to change its style or expression.*), P10 said they would conceal their anger to preserve their relationship with the client:

“...I obviously would be pretty angry, because it’s taking away from my craft and my art in that sense. And it’s a purely financial thing, right? [...] I would probably just say yes, because in the realm of the universe and being grateful, I’m like, well, I’ll be grateful that you hired me in the first place.” - P10

P3 shared that some “actors are able to tell their agents, ‘hey, I do not want to work with this company. I do not stand for what they do,’” but they pointed to the low standard for transparency and control they experience in established industry relationships — leading to an expectation that AI tools might be used on their voice without informed consent. P3 described how VAs often discover their performances have been altered without their input:

“...that actually happens on a regular basis in the industry already, where some of my friends are pitched higher or lower [...] They’ve already hired me and paid me for the job so that kind of already allows them to have permission to manipulate it in a way.” - P3

Beyond their impact on existing labor dynamics, AI corporations have *also* created entirely *new* formats for voice work — often removed from the best practices and protective norms that voice work communities have fought to establish.

VAs have grown suspicious of “audition” calls from AI developers seeking speech data, who may or may not be transparent about their purpose:

“I’ve heard some stories where somebody would audition for a project, never hear from them. And then, when the project came out, it was their voice.” -P3

P11 recounted a recent voice-over client unexpectedly requesting that they submit a video audition—and feeling uneasy upon realizing the client’s intent to collect data on their mouth shape and movement.

“I was asked to do a video audition instead of voiceover... I started really freaking out when I realized that, because they’re not explaining why there are visuals instead for a voiceover, but they’re doing it to take for dubbing projects to compare like, your mouth movement with the mouth movement of the actor, and see how they can manipulate that already[...] that’s using someone’s physicality as well.” -P11

These changes to the landscape of work opportunities disproportionately impact novices like P1, who lack access to the network of union-backed industry roles and established relationships afforded to more experienced VAs. P1 described encountering 3 AI training data job listings - none of which featured fair compensation models – for example, one requested 10 hours of recorded audio for a one-time fee of \$300.

P1 feared that asking for higher compensation from AI companies would be pointless, but expressed empathy for other independent laborers accepting undervalued work:

“Instead of hiring you again, they’ll just go to someone who is like you. who doesn’t value themselves that high[...] People who are just trying to look for the odd job to scrape by[...] are gonna take the AI jobs ... they are being undervalued themselves and being exploited.” -P1

AI developers and researchers hold a responsibility to address these new formats for exploitation that they are enabling and enacting. These disruptions to voice industries stand to disproportionately harm its most vulnerable workers, and degrade labor standards for creative communities at large.

4.2.3 Disrupting the organization and collective negotiation of voice labor. Participants reported that AI rights have become a crucial issue for negotiation in labor disputes that are significantly impacting their careers and lives. This includes the SAG-AFTRA strike (Jul-Nov 2023)[43], the SAG-AFTRA video game strike (ongoing since July 2024) [45], and the ongoing ACTRA labor dispute with Canadian ad agencies (ICA) (ongoing since March 2022). [44]

“There are folks within our community who have fought and worked very hard to create, even during the strike, interim agreements where there’s AI protection in place.” - P14

These ongoing disputes “lock out” union members – preventing them from taking roles in advertising, video games, and interactive media until an agreement is reached.

“We’ve been locked out for over 890 days now. So we’ve got voiceover actors who used to have, like, 6 figure incomes, who’ve lost their houses.” - P13

P11 connected this to their booking rate dropping from 5-6 commercials a year, to 0-1.

“We’ve had a commercial lockout, so they’ve been using non-union performers[...] the amount of opportunities there have been to do commercials are significantly less.” - P11

The established contexts for voice work already place independent creatives in a precarious position; Speech AI’s entry into these spaces exacerbates risks for already disadvantaged creatives.

Involvement in labor communities appeared to have some influence on how voice actors articulated their boundaries and openness toward the use of AI in their work.

Of the 12 “yes” (blue/green, or green) responses recorded (across all participants and scenarios) 10/12 (83%) came from the 5 participants with limited or no reported union involvement (P1, P2, P7, P10, P15). None (0%) of those “yes” responses came from participants who actively participated in labor organization and leadership (P4, P6, P11, P13).

Moreover, the “3 C’s” framework developed by NAVA (the National Association of Voice Actors) was brought up and used in the responses of four separate participants (P4, P6, P9, P11), all active union members. This framework describes Consent, Control, and Compensation as key rights that VAs should ensure are provided for in their work. NAVA distributes this alongside a number of resources intended to support VAs in negotiating their rights, including a list of “Questions to Ask” AI-affiliated clients before agreeing to work with them. [39]

Union resources appear to help voice actors consider, and effectively articulate, their boundaries around creative rights and AI use, preparing them to uphold fair standards for themselves and their community.

5 Discussion

Our study investigated how Speech AI is surfacing as a disruptive force in the present and future of voice work. Our participants aided us in constructing an understanding of the dimensions of labor involved in contemporary voice work, affording a perspective that contextualizes potential and existing disruptions and harms.

We draw from prior work that uses power and socioeconomic structures as a lens for evaluating the impact and role of new technologies in the future of work and creative practice [5, 14, 20, 32] to discuss the implications of our study and provide research and advocacy recommendations for AI researchers, developers, and policy designers.

5.1 Speech AI development trends are reducing, obscuring, and exploiting voice work

The development and deployment of AI systems depends upon obscured human labor, while also frequently and disruptively changing the shape of that labor.

Our findings lead us to consider how the use of voice artists’ performances as training data perpetuates the normalization of

AI “ghost work” [23] — expanding the “hidden layer” of obscured and silenced tech laborers (like dataset annotators) to now, coercively, include the labor of voice artists, too. Even when their contributions are crucial to AI development, these workers are not afforded the same the visibility, compensation, or power as other technology industry workers (like AI developers, designers, and researchers) [23, 25]. This dynamic reenacts Scheuerman et al.’s findings on the widespread devaluation of data labor in dataset research publications [47], and Cheon & Xu’s findings [14] in their investigation of motion-capture animation performers as part of an “creative precariat” — workers prone to the exploitation of their artistry, even as the technology and consumers directly benefit from the uniquely human creative value that they contribute.

Voice artists are also being disrupted by the *deployment* of Speech AI, which we connect to prior studies investigating the ways workers are often forced to take on additional labor when adapting to new forms of work, as shaped by the rapid introduction of automated systems into their industries [5, 20]. Voice workers, their data, their creativity, and their ability to adapt new technologies into their work all represent labor and skills that will continue to be crucial resources to the development of Speech AI systems. In drawing from this prior work, we anticipate that this trend will become increasingly severe as entertainment and service industries strive to effectively apply these technologies in more real-world situations. Yet, voice workers are not being compensated, empowered, or acknowledged on par with the value they contribute. Voice-cloning systems are just one example of an emergent AI technology that is actively exacerbating patterns of exploitation, value extraction, and threats to personal safety [22, 24]. Beyond labor displacement and theft, we found that participants fear abuse of their data for misinformation, deep-fakes, and scams. The same AI misuse risks that threaten the privacy and safety of the general public also threaten voice workers — but the inherent publicity of their voice data, and the asymmetrical power dynamics recurrent in their work, make these performers exceptionally vulnerable.

5.2 Towards power-conscious, community-centered Speech AI development

Drawing from an understanding of voice performance as entangled and embodied creative work, we discuss recommendations for the future of developing Speech AI systems, and other systems that might involve voice artist data collection.

5.2.1 Designing for the reality, not the displacement, of voice work. As in similar labor and AI studies [5, 20], the socioeconomic distance between the disrupted voice workers, and those rapidly developing and deploying AI systems into their work, creates conditions that perpetuate a reductive and disempowering understanding of voice work.

AI developers might begin to tackle this distance through power-conscious participatory design approaches [10] that directly collaborate with voice artists, and focus on developing systems that enable them to *challenge* the unequal power dynamics present in their work [41, 50].

We call on researchers and developers to consider what characterizes speech AI *technologies designed to empower independent creatives*, versus *technologies designed to empower those actively exploiting creatives*; including technologies that can *incidentally* enable disempowerment and misuse.

Both voice workers and clients can stand to benefit from development of new speech technologies; our participants expressed excitement about potential new technologies that could empower their craft and promote accessibility.

We identified how resistant voice artists are to systems automating away the artistic parts of their work that they most enjoy, especially voice performance itself. We also identified the many tasks that voice artists must regularly engage in *beyond* vocal performance, including finding performance opportunities, preparing and submitting audition material, training and rehearsing, negotiating terms and agreements, and participating in community.

Rather than building systems that make voice performance faster or more efficient per se — how might we build systems for the “support tasks” of voice labor, to allow human artists to spend *more* time doing the creative work they *enjoy*? Future work might draw upon creativity support frameworks that outline the ways researchers and system developers can empower creative work at the network or ecosystem level [7, 16, 32].

However, existing development trends are such that powerful industry clients and corporations are positioned to capitalize on the benefits of new technologies first, and without the same risks that individual creatives take on in the face of new technology adoption.

Participants connected their future expectations to their prior experiences with creative labor exploitation; they expressed familiarity with companies prioritizing profit at the expense of human artistry.

Within these structures, **power-conscious Speech AI design can not be achieved through technology solutions alone.** We echo our participants and other researchers and activists in rejecting techno-solutionism, encouraging our communities to focus more resources towards assessing, designing, and upholding ethical standards for working with voice and speech. Our recommendations for future work echo Goetze’s evaluation of the labor theft enacted in the development of Image Generation AI [22] — we must develop and share procedures for empowering data workers and AI research participants, explore new and alternative strategies for ethically sourcing datasets, and develop methods for enforcing ethical expectations in research and industry.

5.3 Towards an equitable future of voice work

While we advocate for involving voice workers in the development of new Speech AI technologies through participatory design approaches, this alone would fall short of actually reworking the existing power structures that exploit the creative precarity of voice artists [4, 14]. Not only do we urge developers to actively center voice actors at each stage of system and dataset development, we also call on developers, researchers, and policymakers to consider what these career performers can teach us about designing, and fighting, for an equitable future of voice work.

Our findings showcase how communities, especially labor communities, provide voice workers access to indispensable support

and collective power. Throughout our conversations, participants pointed us to union standards and resources that equipped them to articulate their rights and boundaries [39, 40]. Participants brought up recent legislative victories in California [1, 2] as some of the most promising developments towards the collective empowerment of actors and performers. Rather than merely pushing for VAs to participate in the technology community, technologists might consider how *they* can meaningfully participate in *worker communities*, and act in solidarity with their organizational efforts.

Future work might include developing support for the labor of creative community organization. How can we help voice workers gain access to these communities and resources, as individuals who are frequently being isolated and shifted across various industries? How might we help support community leaders in the crucial organizational and educational work they do?

Labor rights organizations are doing the work of investigating, organizing, and clearly articulating the demands and needs of performer communities. It is our responsibility to listen and amplify their voices and values in technology spaces, and to directly empower these organizations, honoring their expertise in the design of future systems and policies.

6 Conclusion

It is our hope that, as developers, researchers, and policymakers, we might contribute to amplifying artists' voices, rather than taking them away. Through interviews with 15 voice actors, we have presented a community-centered perspective on Speech AI technology. Our thematic analysis surfaced the sociotechnically and socioeconomically entangled reality of contemporary voice work, and of voice performance itself as an embodied, human craft. We identified the vulnerabilities that these artists face, and the situated expertise that they develop, as they routinely navigate and negotiate their rights within asymmetrical power structures. As Speech AI development drives disruptions that exacerbate existing patterns of exploitation and creative precarity, we examined the crucial role that labor organizations and community networks play, empowering artists in their responses to AI-driven labor disruptions. We discuss the implications for the future of Speech AI development: how directly involving and honoring the expertise of voice workers and community organizers can help lead the development of better technology, and beyond this — how we might commit more efforts towards actively challenging established exploitative power imbalances, to shape a more equitable future for voice work. We call for researchers, developers, and policymakers to design systems that act in solidarity with the community's existing organizational efforts to understand, articulate and defend the needs of voice artists and other workers in creatively precarious positions.

Acknowledgments

We thank ACTRA and NAVA for assisting with recruitment and providing helpful guidance and resources throughout our study. We'd like to thank each one of our participants for taking the time to participate in our study. We'd also like to thank Alex Hanna, Bjoern Hartmann, and our reviewers for their feedback on this paper and throughout our study.

References

- [1] 2024. AB 1836: Use of likeness: digital replica. | Digital Democracy — cal-matters.digitaldemocracy.org. https://calmatters.digitaldemocracy.org/bills/ca_202320240ab1836.
- [2] 2024. AB 2602: Contracts against public policy: personal or professional services: digital replicas. | Digital Democracy — cal-matters.digitaldemocracy.org. https://calmatters.digitaldemocracy.org/bills/ca_202320240ab2602.
- [3] William Agnew, Julia Barnett, Annie Chu, Rachel Hong, Michael Feffer, Robin Netzorg, Harry H. Jiang, Ezra Awumey, and Sauvik Das. 2024. Sound Check: Auditing Audio Datasets. arXiv:2410.13114 [cs.SD] <https://arxiv.org/abs/2410.13114>
- [4] Alex Ahmed. 2020. We Will Not Be Pacified Through Participation. *Tech Otherwise* (oct 15 2020). <https://techotherwise.pubpub.org/pub/qanxilt9>.
- [5] Hunter Akridge, Bonnie Fan, Alice Xiaodi Tang, Chinar Mehta, Nikolas Martelaro, and Sarah E. Fox. 2024. "The bus is nothing without us": Making Visible the Labor of Bus Operators amid the Ongoing Push Towards Transit Automation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 479, 16 pages. doi:10.1145/3613904.3642714
- [6] Ali Alkhatib. 2021. To live in their utopia: Why algorithmic systems create absurd outcomes. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–9.
- [7] Shm Garanganao Almeda, Joy O Kim, and Bjoern Hartmann. 2025. Creativity Supportive Ecosystems: A Framework for Understanding Function and Disruption in Online Art Worlds. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (CHI '25). Association for Computing Machinery, New York, NY, USA, Article 267, 17 pages. doi:10.1145/3706598.3713734
- [8] Julia Barnett. 2023. The Ethical Implications of Generative Audio Models: A Systematic Literature Review. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (AIES '23). ACM, 146–161. doi:10.1145/3600211.3604686
- [9] Sarah Barrington, Romit Barua, Gautham Koorma, and Hany Farid. 2023. Single and Multi-Speaker Cloned Voice Detection: From Perceptual to Learned Features. In *2023 IEEE International Workshop on Information Forensics and Security (WIFS)*. 1–6. doi:10.1109/WIFS58808.2023.10374911
- [10] Susanne Bødker and Morten Kyng. 2018. Participatory Design that Matters—Facing the Big Issues. *ACM Trans. Comput.-Hum. Interact.* 25, 1, Article 4 (Feb. 2018), 31 pages. doi:10.1145/3152421
- [11] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
- [12] Virginia Braun and Victoria Clarke. 2012. *Thematic analysis*. American Psychological Association.
- [13] Simon Caton and Christian Haas. 2024. Fairness in Machine Learning: A Survey. *ACM Comput. Surv.* 56, 7, Article 166 (April 2024), 38 pages. doi:10.1145/3616865
- [14] EunJeong Cheon and Shengyang Xu. 2024. Creative Precarity in Motion: Revealing the Hidden Labor Behind Animating Virtual Characters. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) (DIS '24). Association for Computing Machinery, New York, NY, USA, 3471–3484. doi:10.1145/3643834.3661545
- [15] Cheol Jun Cho, Peter Wu, Tejas S Prabhune, Dhruv Agarwal, and Gopala K Anumanchipalli. 2024. Coding Speech through Vocal Tract Kinematics. *IEEE Journal of Selected Topics in Signal Processing* (2024).
- [16] John Joon Young Chung, Shiqing He, and Eytan Adar. 2022. Artist Support Networks: Implications for Future Creativity Support Tools. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference* (Virtual Event, Australia) (DIS '22). Association for Computing Machinery, New York, NY, USA, 232–246. doi:10.1145/3532106.3533505
- [17] Tom Cole and Marco Gillies. 2022. More than a bit of coding:(un-) Grounded (non-) Theory in HCI. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–11.
- [18] Di Cooke, Abigail Edwards, Sophia Barkoff, and Kathryn Kelly. 2024. As good as a coin toss human detection of ai-generated images, videos, audio, and audiovisual stimuli. *arXiv preprint arXiv:2403.16760* (2024).
- [19] Sefik Emre Eskimez, Xiaofei Wang, Manthan Thakker, Canrun Li, Chung-Hsien Tsai, Zhen Xiao, Hemin Yang, Zirun Zhu, Min Tang, Xu Tan, et al. 2024. E2 tts: Embarrassingly easy fully non-autoregressive zero-shot tts. *arXiv preprint arXiv:2406.18009* (2024).
- [20] Sarah E. Fox, Samantha Shorey, Esther Y. Kang, Dominique Montiel Valle, and Estefania Rodriguez. 2023. Patchwork: The Hidden, Human Labor of AI Integration within Essential Work. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1, Article 81 (April 2023), 20 pages. doi:10.1145/3579514
- [21] Jian Gao and Dashun Wang. 2024. Quantifying the use and potential benefits of artificial intelligence in scientific research. *Nature human behaviour* (2024), 1–12.
- [22] Trystan S. Goetze. 2024. AI Art is Theft: Labour, Extraction, and Exploitation: Or, On the Dangers of Stochastic Pollocks. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (Rio de Janeiro, Brazil) (FAccT '24). Association for Computing Machinery, New York, NY, USA, 186–196. doi:10.

- 1145/3630106.3658898
- [23] Mary L. Gray and Siddharth Suri. 2019. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. HarperCollins. Google-Books-ID: FFJkDwAAQBAJ.
- [24] Wiebke Hutiri, Orestis Papakyriakopoulos, and Alice Xiang. 2024. Not My Voice! A Taxonomy of Ethical and Safety Harms of Speech Generators. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*. ACM, 359–376. doi:10.1145/3630106.3658911
- [25] Lilly Irani. 2015. Justice for “Data Janitors”. <https://www.publicbooks.org/justice-for-data-janitors/>
- [26] Harry H. Jiang, Lauren Brown, Jessica Cheng, Mehtab Khan, Abhishek Gupta, Deja Workman, Alex Hanna, Johnathan Flowers, and Timnit Gebru. 2023. AI Art and its Impact on Artists. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (Montréal, QC, Canada) (*AIES '23*). Association for Computing Machinery, New York, NY, USA, 363–374. doi:10.1145/3600211.3604681
- [27] Zeqian Ju, Yuancheng Wang, Kai Shen, Xu Tan, Detai Xin, Dongchao Yang, Yanqing Liu, Yichong Leng, Kaitao Song, Siliang Tang, et al. 2024. NaturalSpeech 3: Zero-shot speech synthesis with factorized codec and diffusion models. *arXiv preprint arXiv:2403.03100* (2024).
- [28] Haechan Kim, Junho Myung, Seoyoung Kim, Sungpah Lee, Dongyeop Kang, and Juho Kim. 2024. LearnerVoice: A Dataset of Non-Native English Learners’ Spontaneous Speech. *arXiv preprint arXiv:2407.04280* (2024).
- [29] Dennis H Klatt. 1987. Review of text-to-speech conversion for English. *The Journal of the Acoustical Society of America* 82, 3 (1987), 737–793.
- [30] Eun Ju Lee, Clifford Nass, and Scott Brave. 2000. Can computer-generated speech have gender? An experimental test of gender stereotype. In *CHI'00 extended abstracts on Human factors in computing systems*. 289–290.
- [31] Hanlin Li, Nicholas Vincent, Stevie Chancellor, and Brent Hecht. 2023. The Dimensions of Data Labor: A Road Map for Researchers, Activists, and Policymakers to Empower Data Producers. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) (*FAccT '23*). Association for Computing Machinery, New York, NY, USA, 1151–1161. doi:10.1145/3593013.3594070
- [32] Jingyi Li, Eric Rawn, Jacob Ritchie, Jasper Tran O’Leary, and Sean Follmer. 2023. Beyond the Artifact: Power as a Lens for Creativity Support Tools. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) (*UIST '23*). Association for Computing Machinery, New York, NY, USA, Article 47, 15 pages. doi:10.1145/3586183.3606831
- [33] Nan Liu, Shuang Li, Yilun Du, Antonio Torralba, and Joshua B Tenenbaum. 2022. Compositional visual generation with composable diffusion models. In *European Conference on Computer Vision*. Springer, 423–439.
- [34] Xuechen Liu, Xin Wang, Md Sahidullah, Jose Patino, Héctor Delgado, Tomi Kinnunen, Massimiliano Todisco, Junichi Yamagishi, Nicholas Evans, Andreas Nautsch, and Kong Aik Lee. 2023. ASVspoof 2021: Towards Spoofed and Deepfake Speech Detection in the Wild. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 31 (2023), 2507–2522. doi:10.1109/TASLP.2023.3285283
- [35] Yisi Liu, Bohan Yu, Drake Lin, Peter Wu, Cheol Jun Cho, and Gopala Krishna Anumanchipalli. 2024. Fast, High-Quality and Parameter-Efficient Articulatory Synthesis using Differentiable DSP. *arXiv:2409.02451 [eess.AS]* <https://arxiv.org/abs/2409.02451>
- [36] Kimberly T Mai, Sergi Bray, Toby Davies, and Lewis D Griffin. 2023. Warning: Humans cannot reliably detect speech deepfakes. *Plos one* 18, 8 (2023), e0285333.
- [37] Luz Martinez-Lucas, Wei-Cheng Lin, and Carlos Busso. 2024. Analyzing Continuous-Time and Sentence-Level Annotations for Speech Emotion Recognition. *IEEE Transactions on Affective Computing* 15, 3 (2024), 1754–1768. doi:10.1109/TAFFC.2024.3372380
- [38] Rajiv Movva, Sidhika Balachandrar, Kenny Peng, Gabriel Agostini, Nikhil Garg, and Emma Pierson. 2024. Topics, Authors, and Institutions in Large Language Model Research: Trends from 17K arXiv Papers. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*. 1223–1243.
- [39] NAVA. [n. d.]. AI QUESTIONS - NAVA — navavoices.org. <https://navavoices.org/ai-questions/>.
- [40] NAVA. 2023. fAIr Voices: Consent, Control, Compensation — navavoices.org. <https://navavoices.org/fair-voices/>.
- [41] Organizers Of Queerinaï, Anaelia Ovalle, Arjun Subramonian, Ashwin Singh, Claas Voelcker, Danica J Sutherland, Davide Locatelli, Eva Breznik, Filip Klubicka, Hang Yuan, et al. 2023. Queer in AI: A case study in community-led participatory AI. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 1882–1895.
- [42] Vipula Rawte, Amit Sheth, and Amitava Das. 2023. A survey of hallucination in large foundation models. *arXiv preprint arXiv:2309.05922* (2023).
- [43] SAG-AFTRA. 2023. 2023 TV/Theatrical Contracts | SAG-AFTRA — [sagaftra.org. https://www.sagaftra.org/contracts-industry-resources/contracts/2023-tvtheatrical-contracts](https://www.sagaftra.org/contracts-industry-resources/contracts/2023-tvtheatrical-contracts).
- [44] SAG-AFTRA. 2023. ACTRA Needs Our Support — [sagaftra.org. https://www.sagaftra.org/actra-needs-our-support](https://www.sagaftra.org/actra-needs-our-support).
- [45] SAG-AFTRA. 2024. Interactive Media (Video Game) Strike | SAG-AFTRA — [sagaftra.org. https://www.sagaftra.org/contracts-industry-resources/contracts/interactive-media-video-game-strike](https://www.sagaftra.org/contracts-industry-resources/contracts/interactive-media-video-game-strike). [Accessed 22-01-2025].
- [46] Eugenia San Segundo, Paul Foulkes, Peter French, Philip Harrison, Vincent Hughes, and Colleen Kavanagh. 2019. The use of the Vocal Profile Analysis for speaker characterization: Methodological proposals. *Journal of the International Phonetic Association* 49, 3 (2019), 353–380.
- [47] Morgan Klaus Scheuerman, Alex Hanna, and Emily Denton. 2021. Do Datasets Have Politics? Disciplinary Values in Computer Vision Dataset Development. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 317 (Oct. 2021), 37 pages. doi:10.1145/3476058
- [48] Kristen M. Scott, Simone Ashby, and Julian Hanna. 2020. “Human, All Too Human”: NOAA Weather Radio and the Emotional Impact of Synthetic Voices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '20*). Association for Computing Machinery, New York, NY, USA, 1–9. doi:10.1145/3313831.3376338
- [49] Yolande Strengers, Jathan Sadowski, Zhuqing Li, Anna Shimshak, and Florian ‘Floyd’ Mueller. 2021. What Can HCI Learn from Sexual Consent? A Feminist Process of Embodied Consent for Interactions with Emerging Technologies. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (*CHI '21*). Association for Computing Machinery, New York, NY, USA, Article 405, 13 pages. doi:10.1145/3411764.3445107
- [50] Harini Suresh, Rajiv Movva, Amelia Lee Dogan, Rahul Bhargava, Isadora Cruxen, Angeles Martinez Cuba, Guilia Taurino, Wonyoung So, and Catherine D’Ignazio. 2022. Towards Intersectional Feminist and Participatory ML: A Case Study in Supporting Feminicide Counterdata Collection. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) (*FAccT '22*). Association for Computing Machinery, New York, NY, USA, 667–678. <https://doi.org/10.1145/3531146.3533132>
- [51] Harini Suresh, Emily Tseng, Meg Young, Mary Gray, Emma Pierson, and Karen Levy. 2024. Participation in the age of foundation models. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (Rio de Janeiro, Brazil) (*FAccT '24*). Association for Computing Machinery, New York, NY, USA, 1609–1621. doi:10.1145/3630106.3658992
- [52] Vishaal Udandarao, Ameya Prabhu, Adhiraj Ghosh, Yash Sharma, Philip Torr, Adel Bibi, Samuel Albanie, and Matthias Bethge. 2024. No “zero-shot” without exponential data: Pretraining concept frequency determines multimodal model performance. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.
- [53] Apoorv Vyas, Bowen Shi, Matthew Le, Andros Tjandra, Yi-Chiao Wu, Baishan Guo, Jiemin Zhang, Xinyue Zhang, Robert Adkins, William Ngan, et al. 2023. Audiobox: Unified audio generation with natural language prompts. *arXiv preprint arXiv:2312.15821* (2023).
- [54] Yuancheng Wang, Haoyue Zhan, Liwei Liu, Ruihong Zeng, Haotian Guo, Jiachen Zheng, Qiang Zhang, Xueyao Zhang, Shunsi Zhang, and Zhizheng Wu. 2024. Maskgct: Zero-shot text-to-speech with masked generative codec transformer. *arXiv preprint arXiv:2409.00750* (2024).
- [55] Dongchao Yang, Songxiang Liu, Rongjie Huang, Chao Weng, and Helen Meng. 2024. Instructtts: Modelling expressive tts in discrete latent space with natural language style prompt. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* (2024).
- [56] Xueyao Zhang, Liumeng Xue, Yicheng Gu, Yuancheng Wang, Jiaqi Li, Haorui He, Chaoren Wang, Songting Liu, Xi Chen, Junan Zhang, et al. 2023. Amphion: An open-source audio, music and speech generation toolkit. *arXiv preprint arXiv:2312.09911* (2023).
- [57] Xuanru Zhou, Anshul Kashyap, Steve Li, Ayati Sharma, Brittany Morin, David Baquirin, Jet Vonk, Zoe Ezzes, Zachary Miller, Maria Luisa Gorno Tempini, et al. 2024. Yolo-stutter: End-to-end region-wise speech dysfluency detection. *arXiv preprint arXiv:2408.15297* (2024).

A Interview Protocol

Investigating the Impact of Emerging Technology on Voice Actors & Artists: Semi-Structured Interview Guide

(This interview guide has been abbreviated; the version used in practice also included notes and suggestions for wording each question. This guide was developed reflexively, such that the shape of our investigation could co-constructively adapt to the directions suggested by our participants, towards an understanding driven by community perspectives and values.)

A.1 Introduction

- Relay purpose of research

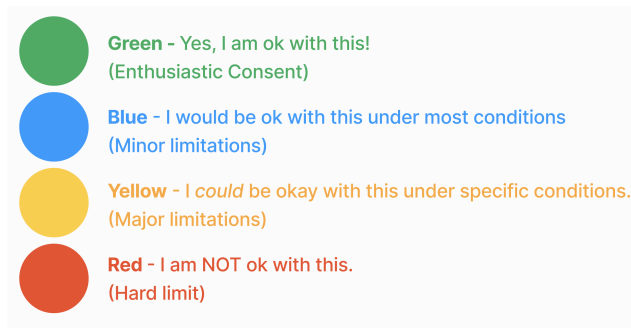


Figure 3: Traffic Light Response Graphic

- Explain format of the interview
- Relay confidentiality
- Ask for permission to record interview
- Tell them recording will start.

A.2 Background Info

(develop rapport, and collect background demographic info)

- Do you identify as a voice artist, voice actor, and/or something else?
- Any relevant adjacent fields (i.e. actor, performer)? What other roles do you identify as?
- What kinds of roles/projects have you worked on?

A.3 Present Perceptions of AI

- Has AI or new and emerging technology come up in your work recently?
- How much experience or knowledge would you say you have with AI and/or the tech industry?
- *(Reflexively added as an official question after this topic surfaced in several interviews):* Are you a member of a larger performance community organization or union?

A.4 Grounding in Existing Research Directions / State of Speech AI Tech with Demos & Samples:

Before each demo, briefly explain what the demonstrated technology is meant to accomplish/ perform .

- AI Voice Cloning
- Text-to-speech voice cloning demo
- AI Voice Editing
- Adversarial (Anti-AI training) Voice Filters
- After each demo, pause and ask for participants response.

A.5 TRAFFIC LIGHTS EXERCISE

Share the Traffic Light Response Graphic with the participant (Figure 3) Participant should choose a color to express how they feel about each scenario.

- Green - Yes, I am okay with this! (enthusiastic consent)
- Blue - I do not mind this.

- Yellow - Uncertain / It depends /I could be okay with this after a discussion.
- Red - Hard limit - I am NOT okay with this.

A.5.1 SCENARIOS: .

- (1) A company wants you to contribute your voice to a speech-generating tool like Siri, or the Text-to-Speech tool on TikTok.
- (2) A company wants to pay you a fee to get your voice data to use in a large training dataset, e.g. to build a model that can give a diverse range of voice performances.
- (3) A company uses your existing voice data (recorded for another purpose) and includes it in a training dataset.
- (4) A research team asks you to contribute your voice to a dataset for a Speech AI research project.
- (5) An team wants to partner with you to help you create your own voice clone product - you would own the clone and control the decision to license it out for companies to use.
- (6) A company wants to edit your voice to change its style / expression

A.5.2 Follow-Up Questions:

- Does compensation matter - what form of compensation would you expect?
- Does the company/team/person's identity matter? (An independent fan, vs. a big tech company?)

What kind of research/technology collaborations might you see yourself participating in, if any? What kinds of conditions would characterize an ideal/respectful cooperation between AI technology and voice actors?

A.6 Wrap Up

- Tell the participant about observations and conclusions from the interview.
- Allow the user to have input into clearing up any misconceptions or to enlarge on important points.

Snowball Recruiting: Can you think of 1-3 other people who you think we could interview? Would you be willing to connect us, or share the screener survey with them ?

Finally, discuss how to send compensation gift card, and thank them for their time.