The Sound of Professionalism: Voice Acting, Technology and Training in Contemporary China

by

Fenglin Selina Ju

A thesis submitted in fulfillment

of the requirements for the degree of

Master's in Chinese Studies

(Master's in International and Regional Studies)

at the University of Michigan

2026

Thesis Committee:

Dr. Jennifer C. Hsieh, Department of Anthropology

Dr. Cara Wallis, Department of Communication and Media

Fenglin Selina Ju

selinaj@umich.edu

© Fenglin Selina Ju 2026

Acknowledgements

I would like to express my deep and heartfelt gratitude to my thesis advisor, Professor Jennifer C. Hsieh, for her patient guidance, enthusiastic encouragement, and valuable critiques of my MA thesis. I'm grateful to Professor Hsieh for being my advisor and for introducing me to key texts in sound studies and a possible approach to pursue for an innovative project beyond academic writing. I also wish to thank Professor Cara Wallis for serving on my thesis committee and for her thoughtful feedback, which prompted me to think more carefully about methodology. This thesis, my intellectual development, and my academic achievements at the University of Michigan would not have been possible without their inspiration and support.

I would also like to thank Professor Ann Lin for her unwavering support in connecting me with available resources. My grateful thanks also extend to Professor Ian Shin, who has always been supportive and encouraging during my PhD application process, job application season, and my exploration of different possibilities at U of M. I also wish to thank Professor S. E. Kile, who has helped me to explore my confusion on methodology and interdisciplinary research, and has guided me to go through the journey at the MIRS program. I sincerely thank Guiqiu Wang, who has made my engagement with the Chinese Studies Center an unforgettable experience, and has provided me with thoughtful feedback on my interview preparation and struggles on the job market. I'm grateful to have Professor Sarah Murray and Professor Diane Oliva as my references for the PhD application, and I'm deeply grateful for their guidance in STS, musicology, and digital media scholarship. I also extend my great gratitude to Dr. Liangyu Fu and Yung-Hui Chou for their continued support and interest in my developing projects, as well as their willingness to share career decisions and

experiences with me. I wish to thank Professor Yucong Hao, my undergraduate advisor, who has been continuously supportive of my intellectual development and personal growth.

This research was supported by the International Institute Global Individual Grant, LRCCS Research – Internship Fellowships, Rackham Conference Travel Grant, and LRCCS Conference Travel Grant, which made my fieldwork in Beijing in the summer of 2025 possible and my conference presentations of the thesis draft over the past year. I am thankful to my colleagues in the MIRS program, the LRCCS cohort, the LRCCS Interdisciplinary Workshop, the "Versatile Bodies in Global Asia" seminar at ACLA 2026, and the Asian Drama and Performance Session at RMMLA 2025 for their feedback on earlier versions of this work in the past two years.

I would like to extend my gratitude to my dear friends, Wenxin Wendy Xiao, Jinting Chrinstine Cao, and Zishan Sherry Qiu, for their companionship, emotional support, and many unforgettable memories we shared that I would like to cherish forever. I'm also thankful to have Chengpeng Li as my friend since the beginning of the program, and I'm grateful for his academic guidance and for being my sounding board. I wish to thank Yanze Zoe Zhu, who makes me realize my potential and resilience as we go through this struggling graduation season. I'm also glad to have Qingyi Zeng, who has always been inspiring, especially during our conversations in Chicago and Spokane. My great thanks also extend to Hange Xie for her advice and for sharing her insights on an idealized lifestyle. I'm also glad to know Gloria Zhou, who recognized my academic research ability and the possible pathways I hadn't realized.

I extend my sincere thanks to the instructors and fellow students at the training program in Beijing, whose generosity and openness made this research possible. This project is indebted to those who welcomed me into their professional world and shared their experiences with me. I do not dare to represent or conclude on the shared experiences and

their professional development. This thesis is dedicated to those who remain committed to their aspiration.

Finally, wish to thank my family—my parents, sister, and partner—for their care and encouragement throughout my years at the University of Michigan. Thank you for always being supportive and helping me every step of the way.

The MIRS program has been important to my pursuit of a PhD career. However, life is always full of challenges and unexpected circumstances. In the past three months, I've decided to redirect my aspirations and career. It is not an easy decision to make, since working towards a doctoral degree has been my goal, at least for the past three to four years: since my work originating from personal life was awarded, since my projects attracted audiences with different backgrounds, and since my dedication to join the University of Michigan. What I've been trained and developed, my critical thinking, my ability to conceptualize research, my sensitivity to scholarship, were not granted with acknowledgment as I initially expected or dreamed about. Fortunately, I'm still the one who is always curious about possibilities, who dares to take risks, and who seeks to grab opportunities. It's been a painful, struggling, and challenging journey. But I never regret any of the decisions I've made.

Table of Contents

Acknowledgements ..... **i**
Abstract ..... **iii**
Chapter 1 Introduction **1**
The Genealogy of Chinese Voice Acting 2
Research Question 5
Fieldsite and Method 9
Bridging Sound, Matter, and Labor 14
Chapter 2 Acoustic Infrastructure **19**
Studio and Infrastructure 21
Media Infrastructure, Sonic Infrastructure, and Acoustic Infrastructure 23
The Material/Acoustic Dimension: The Acoustic Space 27
The Institutional Dimension: For the First Time 31
The Relational Dimension: When the Infrastructure Breaks Down 34
Coda 36
Chapter 3 The Technovocalic Body **39**
The Voice Behind the Screen 41
Technovocalic Body/Technovocality 43
The Body Full of Memory 46
How to Catch the Digital Frankenstein 52
Conclusion: A Multimodal Future **55**
Appendix **61**
Bibliography **62**

Abstract

In light of the emerging auditory pop culture of the 2010s, voice acting (*peiyin* 配音) has increasingly been seen as a promising career path and a shared interest among online communities in China, particularly among enthusiasts of Internet radio dramas, anime, and video games. Intersecting the legacies of the film industry, TV production, and Japanese seiyu (voice actor) culture, the Chinese voice acting industry is dominated and professionalized by voice agencies and studios whose core businesses span these audiovisual formats.

Focusing on the recording studio as a key site of institutional intervention and production environment, this thesis explores what constitutes the "sound of professionalism" in contemporary Chinese voice acting and how we might reconceptualize the dialectic between bodily materiality and technological mediation in the situated practice of voice acting. Building on recent scholarship that shifts the conceptualization of voice from an expression of natural interiority to its material aspects, I argue that voice is co-constituted through bodily actions and technological mediation as a body-technology interface. By introducing and theorizing the framework of acoustic infrastructure and technovocalic body/technovocality, I describe how vocal performance emerges through the interaction between bodily training and recording technologies. In doing so, this research challenges technological determinism and the flat ontology assumptions present in the New Materialism that assigns equal agency to both human and nonhuman actants.

Drawing on ethnographic fieldwork I conducted in the summer of 2025 at a voice acting training program hosted by a professional voice agency in Beijing, I suggest that, in this context, practitioners' limited understanding of technological affordances and what sound actually is mediates their relationship with the production environment, prompting them to adapt their bodies to meet opaque technical parameters. In other words,

professionalism in the Chinese voice acting industry emerges from an asymmetrical power distribution between practitioners and the technological infrastructure. By foregrounding the lived, embodied experiences of voice acting, this research contributes to ongoing conversations in Science and Technology Studies (STS), sound studies, and New Materialism while bringing attention to an understudied domain of contemporary Chinese popular culture.

Keywords: technovocalic body/technovocality, acoustic infrastructure, materialist turn, professionalism, voice acting, Chinese popular culture

Chapter 1 Introduction

In January 2018, Hunan TV aired a special TV show targeting fans fascinated by attractive, expressive voices. Dedicated to featuring the charisma of voice acting, this show, *The Sound*, invited well-known actors to dub classic films, TV dramas, and anime clips, and finally selected an outstanding performer crowned as the "King of Voice." As the first entertainment TV show to center on voice performance, *The Sound* quickly became the most-watched program in its time slot, drawing public attention to voice acting, a skill that often remains behind the scenes of audiovisual production. In December 2020, the video platform Bilibili launched *Voice Monster*, which features competitions and performances by professional voice actors. Unlike the 2018 show, which primarily invited active screen actors, *Voice Monster* showcased full-time voice acting practitioners from diverse backgrounds beyond performing arts training. With millions of heated online discussions, these shows have introduced voice acting not only as an entertaining spectacle but also as a promising career path. These developments reflect the rapid expansion of China's so-called "ear economy (*erduo jingji* 耳朵经济)," which, according to Ying Huang, **is "a form of economy that uses hearing as the main information acquisition method and webcasting, audiobooks and knowledge payment as the main content" and is seen as the competent and an important collaborator of screen culture in the future.**[1] As of 2024, approximately 700 million consumers seek high-quality online entertainment in audio-based or audio-enhanced formats, which Huang identifies as evidence of the ear economy's rise.[2] Industry reports further estimate that China's voice-over industry, valued at 3.704 billion RMB in 2018, is expected to reach approximately 16 billion RMB by 2030. While the growing ear economy offers unprecedented visibility and commercial appeal, voice acting did

[1] Ying Huang, *Ear Economy: China Audio Streaming Programs Study* (Singapore: Palgrave Macmillan, 2024),136, 226.
[2] *Ibid.,* 109.

not emerge in a vacuum and has a much longer, more complex historical legacy in Chinese media industries. The current popularity of voice-centered entertainment is rooted in earlier institutional and technological developments, particularly the practices of film and documentary dubbing, television production, and transnational influences from the Japanese *seiyu* (voice actor) culture.

The Genealogy of Chinese Voice Acting

Voice acting has never been an isolated or autonomous performing genre. Instead, as the performance clips featured in the aforementioned shows suggest, it has historically emerged at the intersection of different media infrastructures, including film dubbing, television post-production, and transnational animation industries. Before tracing this history, it is necessary to clarify the terminology surrounding voice acting. So far, I see the phrase "voice acting" as the English equivalent of "*peiyin* 配音," which literally means "providing something with sound." According to the Xinhua Dictionary, *peiyin* is the voice work that translates foreign films or replaces poor on-site recordings with dialogue, narration, sound effects, and music in film production, which strongly aligns with the definition of "dubbing" in the Oxford Languages Dictionary. This interpretation dictates the usage of *peiyin* as dubbing in cinematic works, especially in the 1950s PRC.

Dubbing aims to provide translations for foreign films, most of which came from countries within the newly friendly socialist bloc, especially from the Soviet Union, with a steadily shrinking number of US-imported films.[3] Beyond removing the language barrier, translated films were expected to fulfill an educational function and remain accessible to workers, peasants, and soldiers. The dubbing work was expected to maintain a balance between Sinicized translation (*zhongguo hua* 中国化) that does not simply apply Chinese

[3] Lin Feng, "Voicing the West: Dubbing Artists and Invisible Labour Behind the Socialist Screen," in *Film History and Screen Culture in and beyond Greater China*, ed. Lin Feng (London: Routledge, 2025), 74, https://doi-org.proxy.lib.umich.edu/10.4324/9781003378839.

linguistic conventions or regional dialects wholesale, but also considers the original context of the film with understandable expressions, and artistic creation, which unravels the character's personality, emotions, and ideological commitments.[4] In other words, a successful dubbing involves not merely lip-synchronization but also appropriate acting with vocal skills. Therefore, the art of *peiyin* carries on the cinematic aesthetic tradition. In this context, voice acting functioned primarily as a translational and ideological practice embedded within film production.

Similarly, voice acting in the 1990s can also be understood as dubbing, given its ability to complement underdeveloped recording technologies in TV production. According to Xia Lei, one of the popular voice actors with 16 years of experience, voice acting is an inevitable part of post-production, which functions as Automated Dialogue Replacement (ADR) to replace unusable on-site audio caused by environmental noise, microphone issues, or collaborations with actors speaking different languages.[5] Rather than inviting screen actors themselves to record lines, recruiting independent voice actors became a more convenient and cost-effective solution. In addition, the pedagogical limitations of theatre academies make post-production voice acting both necessary and inseparable. Current acting trainings, such as in Beijing Film Academy and Shanghai Theatre Academy, are divided into four individual disciplines: voice, speech, acting, and movement (*sheng tai xing biao* 声台行表), leaving students with the burden to integrate them together by themselves organically.[6] As screen actors focus more on their acting, with poor vocal performance, voice actors are responsible for lip-syncing and improving emotional delivery, thus addressing the limitations of TV recordings.

---

[4] Zheng Xiaofeng. "One Year of Soviet Film Translation Work: Several Issues in the Artistic Creation of Dubbed Films 一年来苏联影片的译制工作:翻译片在艺术创造上的几个问题 yinian lai sulian yingpian de yizhi gongzuo: fanyi pian zai yishu chuangzao shang de ji ge wenti." *Guangming Daily*, March 17, 1951.
[5] Xia Lei 夏磊, "走进 '替'声的世界, 讲述真实的配音行业," *TED*中文, 2016, https://www.ted.com/podcasts/tedinchinese-transcripts-xialei2016.
[6] Warren Chu, "Bridging Linklater Voice and Meisner Acting within the Chinese Classroom," *Voice and Speech Review* 19, no. 1 (2022): 6 https://doi.org/10.1080/23268263.2022.2143604.

What the following chapters understand by voice acting is not limited to voice work in film and TV production and is largely influenced by Japanese seiyū (voice actor) culture. Since the rise of the anime industry in the 1970s, especially the limited animation technique, which reuses frames and animates fewer parts, Japanese anime opened up space for seiyū to voice the reduced visuals. To advertise their work to the public, seiyū were affiliated with production companies and promoted as icons. Particularly after the 1990s, they have been idolized as popular stars, hosting public performances beyond mere studio recordings, including singing, podcasting, live performances, and idol group activities. Unlike backstage practitioners, they are highly recognized for their efforts and talents, which provide voices or "ensoul" characters for anime, video games, and radio dramas. In other words, voice acting does not merely replace on-site poor performance with new recordings of lines, but makes the actors themselves the medium to transmit characters' voices to the audience.[7] In China in the 2000s, voice acting emerged as an online practice among fans of Japanese anime and radio dramas. In the early stages, Chinese voice actors were independent practitioners without the academy-based training that Japanese seiyū had. They established online interest groups to produce non-commercial Internet radio dramas for free publication and later adopted a broader range of works, including fan fiction, manga, video games, and anime.[8] As Xia Lei mentioned, although film and television dubbing dominated their opportunities, it was only after the emergence of the Chinese anime and video game industry that voice acting was "professionalized." Popular online actors, respectively, organized commercial voice agencies or studios, directing Chinese voice acting along the Japanese model through institutional management. As voice acting is not an individual discipline in art academies but is often affiliated with programs such as broadcasting, theatre, and performance, the institutional

[7] Shunsuke Nozawa, "Ensoulment and Effacement in Japanese Voice Acting," in *Media Convergence in Japan*, ed. Patrick W. Galbraith and Jason G. Karlin (Ann Arbor: Kinema Club, 2016), 187.
[8] Tingting Hu, Jing Jin, and Lin Liao, "A Transmedia 'Third' Space: The Counterculture of Chinese Boys' Love Audio Dramas," *Asian Studies Review* 47, no. 4 (2023): 838, https://doi.org/10.1080/10357823.2023.2211223.

support at the academy level cannot meet young practitioners' expectations. Therefore, the independent voice agencies and studios become the dream workplaces for those starting their careers through online practices.

Research Question

Existing scholarship focuses on industry growth, fan cultures, or media circulation, leaving the production conditions underexamined. For instance, voice acting (or sometimes used as a synonym for dubbing) often serves as a powerful supplementary material for analyzing the aesthetics, sensory value, and embedded feminist implications of popular audio products. Among cases focusing on the adaptation of Internet radio drama from text and the sonic design in Otome games, voice acting is emphasized for voice's ability and texture that stimulate a sense of intimacy and presence beyond mere semantics, such as different vocal timbres as the central identifier of characters and whispers or gasping sounds representing the intimate atmosphere.[9] However, the material conditions of these auditory products are rarely examined. The voice is treated as an innate release of charisma for close listeners, which shadows the production and mediation before it reaches the audience's ears. This tendency is partially corrected by scholarship that takes listening technologies seriously as mediators rather than neutral conduits. For example, Jasmine Yu-Hsing Chen traced how personal audio devices, such as wired headphones and wireless headphones, have transformed audio content from a collective, politically mobilizing medium into an intimate, individualistic one, using Chinese Internet radio dramas as an example. Voices are transmitted directly into listeners' heads, intensifying the affective charge and the privacy they create.[10] However, the analytical

[9] Tingting Liu and Zishan Lai, "Eroticism and Design Workshops in Game Interfaces: Exploring a Mobile Otome Game for Chinese Women," Games and Culture, December 22, 2024, 11-12, https://doi.org/10.1177/15554120241306788.; Yucong Hao, "Transmedia Adaptation, Sonic Affect, and Multisensory Participation in Contemporary Chinese Danmei Radio Drama," *Global Storytelling: Journal of Digital and Moving Images* 3, no. 1 (July 31, 2023): 75, https://doi.org/10.3998/gs.3771.
[10] Jasmine Yu-Hsing Chen, "Beyond Words: Navigating Aural Fantasies in Chinese Boys' Love Audio Drama," *M/C Journal* 27, no. 2 (2024): [specific page], https://doi.org/10.5204/mcj.3033.

lens remains firmly on the receiving end, where mediation is theorized as something that happens to listeners.

This is not to say that technological mediation in audio production has gone entirely unnoticed. Some scholarship has already addressed the significance of technological mediation, such as the positioning of different microphones and standardized effects that represent certain objects and actions, in elevating sound with expected fantasy. Yet, the discussion has empowered microphones as the more important production technique than actors themselves, leaving little space to acknowledge actors' bodily negotiation.[11] In other words, practitioners who produce that fantastical voice, the studio environment, other than acoustically complex design within which that voice is made, and most importantly, their underlying interactions are outside the frame. The everyday contingent process by which practitioners are situated within production environments remains largely opaque.

To unpack and recognize the importance of everyday encounters, I see analysis of the studio as the primary approach. Studio studies have generated rich accounts of how spatial design, equipment, and acoustic treatment shape musical performance and production. In Susan Schmidt Horning's research on the history and transformation of sound recording technology, she explained that when highly sensitive microphones were first introduced into recording sessions, recording experts were responsible for controlling and adjusting the room acoustics to accommodate different instruments and scores.[12] Here, the studio, which contains advanced devices and systems, and its space collectively contribute to and construct the in-room music performance. Though voice acting studios share certain acoustic logics with music recording environments, as Horning demonstrates, including space and equipment collectively shape performance, studio studies for non-music use and how practitioners in this

[11] Neil Verma, *Theater of the Mind: Imagination, Aesthetics, and American Radio Drama* (Chicago: University of Chicago Press, 2012), 57.
[12] Susan Schmidt Horning, "The Studio Electrifies: Radio, Recording, and the Birth of the Small Studio Business," in *Chasing Sound: Technology, Culture, and the Art of Studio Recording from Edison to the LP* (Baltimore: Johns Hopkins University Press, 2013), 42.

field understand and collaborate with the environment have received little attention. In the Chinese context, where voice acting serves multiple industries simultaneously, including film dubbing, animation, games, and Internet radio drama, this studio environment becomes a particularly charged site where institutional standards, bodily training, and technological parameters converge. It is here that aspiring voice actors transition from online hobbyists to studio practitioners, navigating not only artistic performances but also the opaque technical conditions of professional audio production. It is in this studio environment, where bodily training meets technological infrastructure, that the question of professionalism becomes most visible and most contested. With the studio as the dominant production setting, voice acting is neither purely a technical procedure nor simply an artistic performance. Rather, it emerges from the interaction between the trained body and recording technologies. This paper, therefore, asks: What constitutes the "sound of professionalism" in contemporary Chinese voice acting, and what material and technological conditions make it possible? More specifically, how does vocal professionalism emerge from the interaction between bodily materiality and technological mediation in the situated practice of voice acting?

This research largely draws from Jonathan Sterne's statement regarding how sound is inherently mediated, where the techniques of audio media, such as MP3, condition and shape the listening environment and "gives the full experience of listening to a recording while only offering a fraction of the information and allowing listeners' bodies to do the rest of the work."[13]I argue that the relationship between the recording studio and its practitioners works in a similar way. As Sterne suggests, MP3s require bodily techniques, including anticipating the inner ear's movements to modify how sound is perceived, even though they are not bodily techniques themselves but rather "model reception to affect it."[14] Similarly, the practice of voice acting also requires a concordance of the studio environment as an active mediator

[13] Jonathan Sterne, "The MP3 as a Cultural Artifact," *New Media & Society* 8, no. 5 (November 2006): 834-36, https://doi.org/10.1177/1461444806067737.
[14] Sterne, 837.

rather than a transparent medium, and the bodily actions, where practitioners train themselves to fit their voices in.

Specifically, the body-technology interface engendered from the practice presents an asymmetrical distribution of power and subjectivity. In other words, the recording studio, or what I term the acoustic infrastructure of voice acting, is recognized for its efficacy but is not accorded agency, a term closely associated with human subjectivities. Humans are more responsible parties with greater capacity to interpret, reconfigure, and create meanings that extend beyond the intentional use of technology. Yet, the effects of technology remain uncertain and can arise from unexpected, creative interactions with its users and the underlying social politics.[15] Hence, in the materialized action of the voice acting practice, my research neither authorizes recording technology with absolute power to determine production nor reduces it to a passive tool. Instead, this research will analyze the overlooked importance of the studio and the take-for-granted voice performance. I argue that limited exposure to the studio, as well as performances and habits that carry on the legacies of voice training in both art academies' pedagogy and contingent body adjustment, contribute to the institutional professionalization of voice acting when voice studios dominate the industry. As voice actors are only required to be responsible for their performance, the clarity, charisma, and emotional delivery of the voices, their encounters with the studio and the whole production environment are carefully managed. At the same time, the institutional power of the space mediates how they come to understand themselves becoming professionals. The process of professionalization reveals the production of the technovocalic body, one that is "produced specifically through media technologies and their reconfiguration of the relationship between recorded body and voice."[16] Unlike seeing voice as a natural process

[15] Jenny L. Davis, "Politics and Power," in *How Artifacts Afford: The Power and Politics of Everyday Things* (Cambridge, MA: MIT Press, 2020), 57-58, https://doi.org/10.7551/mitpress/11967.003.0005.
[16] Jaimie Baron, "Ventriloquizing Obama, or, the Ethics of Archival Ventriloquism," in *Media Ventriloquism: How Audiovisual Technologies Transform the Voice-Body Relationship*, ed. Jaimie Baron, Jennifer Fleeger, and Shannon Wong Lerner (Oxford: Oxford University Press, 2021), 200.

arising from interiority and voice acting as a natural talent, technovocality describes how vocal acting emerges from and is shaped by the practitioner's interaction with recording technologies, and theorizes how the body learns to perform within specific technological constraints whose full logic remains opaque to the performer. Learning from instructors with art academy backgrounds at a popular voice studio, though not with a voice acting major, contemporary voice acting has a historical continuity of broadcasting training. Simultaneously, the "blackboxness" of the recording studio becomes constitutive of technovocality itself, prompting the actors to unconsciously adjust their bodies and voices to fit the environment. Based on the underlying process, the acting program that happened in the recording studio can be theorized as a "materialized action" according to Jenny Davis, "the material manifestation of human values and biases."[17]

Fieldsite and Method

After entering the gate and turning left, one descends a short flight of stairs into the recording studio, a walkout basement with a public space inside the building and a soundproofed room in the corner. This is where practitioners usually work and where students have in-studio lessons during the three-week program. As part of a program series launched by Manzhi MZ Studio across summer and winter terms in June 2023, this 9th program began in mid-July 2025 in Beijing, China. Affiliated with Manbo, one of the most popular audio content platforms developed and operated by Beijing Yuca Interactive Technology Co., MZ Studio introduced online courses and short-term in-person training programs for young fans of Internet radio drama and voice acting. Compared to other training programs from famous labels, such as Voice 729 and Bianjiang Studio, which implement merit-based selection for future colleagues, the MZ program offers a rare entry point for "outsiders" into the voiceover industry with minimal selection criteria, notably the ability to speak standardized Chinese

[17] Davis, 57.

(*Putonghua*). However, a relatively high program fee sets another barrier. Besides guest speakers from the above-mentioned famous labels in some sessions, another attraction of the program is the opportunity for students to practice and receive feedback on their performance, from film to Internet radio drama, inside the studio.

The program lasts for 21 days with an average of 8-9 hours per day. Except for the first two days, when students practice basic performing skills in the auditorium, the rest of the program is hosted in the recording studio. Spanning topics across modern TV dramas, costume dramas, cartoons, anime, video games, and Internet radio dramas, students either pick up their own pieces or are assigned characters by the instructor. They take turns practicing and recording inside the booth and receive feedback from the instructor. By the end of the program, students will present their work in the large classroom, where the graduation is announced. I was one of the 17 students, most of whom majored in broadcasting and performing arts in college and were in their early 20s. After watching *Voice Monster* and engaging with popular culture products featuring famous voice actors, the other 16 students, except for me, aim to pursue careers as voice actors, and some already have experience in Internet radio drama production. Art college programs in China do not offer voice acting as an independent major. Instead, it is usually affiliated with programs such as broadcasting or, more often, a general elective course. Although they received similar performance training in college, such as line practice, stage performance, and speaking, the students found that their academic education did not prepare them for the dream job. Thus, having training experience at professional voice agencies and studios, with instructors and guest speakers from major labels, is highly desirable.

This research draws on field notes, performance recordings, and embodied perceptual memories accumulated through my participation in the program. In other words, this research is deeply grounded in the situated knowledge that emerges through practices. As Ingold

argues, the form a practitioner produces does not pre-exist as a blueprint in the mind to be mechanically transcribed onto a neutral medium. Rather, it "comes into being through the gradual unfolding of that field of forces set up through the active and sensuous engagement of practitioner and material," cutting across the emergent interface between them.[18] Skill, in this sense, cannot be adequately captured through retrospective accounts. What voice actors do in the recording booth is not the outward expression of an interior vocal idea but an emergent negotiation shaped by the parameters of the environment they inhabit. As he further suggests, technical constraints do not prefigure the form that emerges but "set the parameters of the process" within which it unfolds.[19] Thus, the contingent, emergent process of becoming professional must be observed as it unfolds. This is precisely what sensory ethnography makes possible. Rather than treating sound as the sole object of inquiry, Sarah Pink's framework attends to the full multisensory field through which knowledge is produced in practice. By trying hard to perform and carefully listening to my and others' bodily actions, the particular sound-oriented research responds to what Pink contested as the visual-centric ethnographic experience through participating in the career of becoming a voice actor, thus attending to the continually constituted place and senses, and the situatedness of myself as the ethnographer.[20]

By exploring my situated knowledge and experience, I claim this research as an autoethnographic project. The use of autoethnography here is not intended to make a distinction from ethnography. As Heewon Chang explains, autoethnography "follow(s) a similar ethnographic research process by systematically collecting data, 'field texts' in Clandinin and Connelly's (2000) words, analyzing and interpreting them, and producing scholarly reports, also called autoethnography. In this sense, the term 'autoethnography'

---

[18] Tim Ingold, "On Weaving a Basket," in *The Perception of the Environment: Essays on Livelihood, Dwelling and Skill* (London: Routledge, 2021), 430, https://doi.org/10.4324/9781003196662.
[19] *Ibid.*, 434.
[20] Sarah Pink, *Doing Sensory Ethnography*, 2nd ed. (London: SAGE Publications, 2015), 31, 33, https://doi.org/10.4135/9781473917057

refers to the process and the product, just as 'ethnography' does."[21] What is different about autoethnography is the use of personal experiences as primary data, not only "about focusing on the self alone, but about searching for understanding of others (culture/society) through self."[22] In other words, the autobiographical narratives are not evaluated for their inherent biases, but are valued for the efforts to connect the self with others, thus allowing "the stories of autoethnographers to be reflected upon, analyzed, and interpreted within their broader sociocultural context."[23]As a participant and researcher, my embodied experience lies not only in the analysis of my personal interpretation or solely in others' insights. Instead, I actively learned, practiced, and participated in the program as others did, and also included my informants' self-narratives. The central arguments of this thesis, that the studio's acoustic infrastructure shapes voice acting production in ways that remain opaque to practitioners, and that the technovocalic body is formed through embodied negotiation with technological parameters, depend on a kind of evidence that only becomes available through my own bodily experience, which also extends beyond the individual. Thus, participant observation is an important component of my method, through which I collect information by observing others' interactions during the auditing sessions. Rather than a method, autoethnography serves as the leading methodological framework for this research, especially in moments when my embodied experiences are more evident materials supporting my arguments.

My particular position within the field inevitably shapes this multisensory research. Unlike the other sixteen students who entered the program with the explicit goal of pursuing voice acting as a career, I participated as both a learner and a researcher. This difference was not concealed. I disclosed my research intentions to the instructor before the program began and obtained verbal consent from participants. All student participants are referred to by pseudonyms, with identifying details adjusted where necessary. As a learner, I was required

[21] Heewon Chang, *Autoethnography as Method* (London and New York: Routledge, 2016), 48.
[22] *Ibid.*, 48-49.
[23] *Ibid.*, 46.

to practice, record, and receive feedback alongside others, enabling what Rod Davies describes as a combination of etic and emic perspectives — the researcher's analytical distance alongside an insider's embodied participation."[24] This shared experience grounded my observations in embodied participation rather than detached scrutiny. My positionality and the biases it carries are thus inseparable from the research process. This participation in the educational setting constitutes an "apprenticeship of listening" in learning to hear and respond as my informants and widely anticipated professionals do.[25] The self and autoethnographic reflection are not a limitation but a condition of the research object itself. As voice acting in China lacks a standardized academic training pathway, my fieldwork with my presence, reactions, and responses in this beginner-level program precisely contributes to and demonstrates that the transition to expected professionals occurs through the training process, the moment of its formation.

At the same time, the nature of the data, consisting of field notes written during and immediately after sessions, recordings of my own performances, and embodied perceptual memories accumulated through repeated participation, requires acknowledgment. Much of what this thesis analyzes—micro-adjustments of posture and breath, real-time sensory feedback through monitoring headphones, comparative observations when the same piece was recorded under different conditions—cannot be recovered from audio recordings alone. Since these embodied, technologically mediated dimensions of vocal practice resist capture through conventional recording and interview methods, this research requires an epistemological orientation that treats sound itself as a modality of knowing rather than merely an object to be documented. By listening to and participating in sonic events that are not constituted solely by audible objects, I engage in what Steven Feld terms acoustemology:

---

[24] Rod Davies, "An Autoethnography of Performance in the Vocal Booth," in *The Routledge Companion to Music, Autoethnography, and Reflexivity, ed. Peter Gouzouasis and Christopher Wiley* (London: Routledge, 2024), 89.
[25] Tom Rice, "Ethnographies of Sound," In *The Routledge Companion to Sound Studies*, edited by Michael Bull (London: Routledge, 2018), 243-45.

an exploration of sonic sensibilities as a way of knowing and being in the world. Acoustemology, as he develops it, is not simply a method of attending to sound but an epistemological orientation in which sound serves as a primary modality of knowing, "one does not simply 'acquire' knowledge but, rather, that one knows through an ongoing cumulative and interactive process of participation and reflection."[26] Over the 21 days of training, my ears were progressively attuned to distinctions I could not initially perceive, such as the difference between my voice as I heard it through bone conduction and as it was returned to me through monitoring headphones, the subtle changes in tonal quality when I adjusted my distance from the microphone or when others' adjustments heard by me, and the gap between what I thought I had corrected in a subsequent take and what the director still heard as unchanged. Similar to how Feld reflected on his listening and on others' listening in the forest and in recordings, the world of the studio and Chinese voice acting is constituted relationally through the conjunctions, disjunctions, and entanglements of human, non-human, material, and technological elements.[27] Hence, the relational situations for voice acting demand a studio acoustemology to revisit and then theorize the ongoing, contingent processes of becoming professional.

Bridging Sound, Matter, and Labor

When asked "what attracted you to join the summer program," some, with previous experiences of learning voice acting in college, responded as follows:

"In terms of the workplace, the quality of the school's studio is very poor, and you don't have the chance to record in the studio every day. Doing voice acting outside (of a professional studio), you don't even want to record, from the environment of the room to the quality of the sound. Also, the teachers at school instructing voice acting are acting-line specialists, and they don't know much about voice acting at all."

[26] Steven Feld, "Acoustemology," in *Keywords in Sound*, ed. David Novak and Matt Sakakeeny (Durham, NC: Duke University Press, 2015), 13-14.
[27] *Ibid.*, 12, 19.

One student majoring in broadcasting further explained her opinion of the school-outside program comparison:

"In the broadcasting major, it's fair to say that no teacher understands what voice acting is, as they are not really voice-acting practitioners in the first place; they are looking at their own experience of broadcasting to judge the performances. For example, one of my teachers focuses on how the voice-over corresponds to the mouth shape in the picture, with the emotional performance secondary. Indeed, you can understand that school teachers are weaker in this aspect of voice acting; it's not their strength. They are better at telling you what voice acting is in theory, but in practice, it is a different concept from what we understand. Studying in the studio, especially during the days when teachers from 729 (one of the most famous voice acting agencies in China) were in class, I personally felt a big difference!"

From the students' responses, it was clear that there were two main reasons why they came to the program: 1. a more professional studio environment compared to school or their own homes; and 2. a more hands-on and perceived "professional" voice acting training, as opposed to the more theoretical teaching at school. It was conversations like these that got me interested in the question of "what is the professional sound." It was also my peers' responses that led me to organize this research into two components: the production environment, namely the studio, and bodily training.

In the following chapters, I will elaborate on the concepts I mentioned above through rich ethnographic cases and reflections: acoustic infrastructure and the technovocalic body. In Chapter 2, I will explain the criteria for infrastructure and how the recording studio is counted. By deconstructing categories of infrastructure, such as media and sonic infrastructure, this chapter will focus on how acoustic infrastructure and its spatial, material, and institutional attributes are featured and revealed through the case of in-studio practices. Carefully documenting and translating the ethnographic experiences, my intervention with the acoustic infrastructure centers on three interwoven perspectives: the studio as a material/acoustic, institutional, and relational space. By inviting cases of how career-starter voice actors are expected and exposed to default industry standards, ranging from working

routines and behaviors to signal processing, the studio as acoustic infrastructure operates through interlocking dynamics that collectively produce a partially opaque professional environment. This environment exercises institutional power by maintaining its invisibility and conforming to rules. In other words, when everything is functioning properly, practitioners perceive the professional standard as a neutral measure of their own skill and preparation, without realizing that this standard itself has been produced by the acoustic infrastructure they inhabit, through arrangements of space, technology, and institutional routine that remain largely outside their awareness.

Chapter 3 begins by departing from the illusion of audio-visual synchronization and turns to the production process of creating the illusion. Situating voice acting as a body-technology interface and individual embodied memory, rather than examining the social relations behind the scenes, this chapter analyzes how voice is trained through repeated practices and training materials, including academic textbooks and inherited formulaic performance styles and methods. When practitioners enter the acoustic infrastructure, what happens to them is not just the acquisition of technical skills but the reconfiguration of the body itself, a process that eventually occurs under partially opacified conditions. By introducing "technovocalic body/technovocality" from previous studies of digital ventriloquism and technologized vocal training that contribute to the production of AI voices, I unpack how the technologization of voices is being internalized through unconscious reflection and naturalized through technical processing. Thus, the process of professionalism formation is a gradual descent from deliberate training to involuntary physical and technical intra-action, ultimately being institutionalized without the full knowing of the practitioner.

The conclusion chapter is a new attempt. Instead of concluding my arguments and potential contributions in the same writing style, I decided to take on a multimodal approach. As my analysis is mostly based on sensory experiences, including listening, observing, and

other possible bodily engagement, the cases I introduce in this research are not direct recordings with semantic meaning. A methodological question comes to the surface: what are the possible ways to preserve, document, and present sound in an inherently static, structured academic writing? Thus, in this chapter, I show how drawings can be a creative and critical way to visualize my irreplicable fieldwork.

By theorizing what I term "acoustic infrastructure" and the development of technovocality in this context, my research aims to interpret the sound of professionalism in Chinese voice acting. Professionalism is defined not by mastery of specific technological knowledge but by adaptation and negotiation between material environments and bodily labor.

As I introduce how career starters learn to become professionals, this research reveals how professionalism itself is constructed and legitimized through asymmetrical encounters between bodies and technologies. In industries where technical infrastructure is invisible to practitioners, the standards that infrastructure imposes become attributed to individual talent and skills or institutional training, obscuring the material conditions that produce them. By integrating a discussion of technovocality, this study rejects the separation between expertise in the body and in technical knowledge. It names the process by which the body is trained not in abstract space but within specific technological apparatus, and it draws attention to what happens when those constraints remain unacknowledged. In a broader sense, this research speaks with scholarship responding to the general claims, "things have politics," that many new materialists insist on. Inspired by Abrahamsson et al.'s discussion of omega-3 as "matter related" instead of "matter itself," which largely develops from and further expands Jane Bennett's research on omega-3's agency and causal effects, the sound of professionalism in voice acting is not matter itself either, but is constitutive of conditions, the environment

designed to produce it, and the contingent human participation.[28] Similar to what they argue about omega-3, saying "omega-3 is never alone. Whatever it does, it does in conjunction with other entities, enmeshed in a variety of relations," so does the production of professional sound in voice acting.[29]

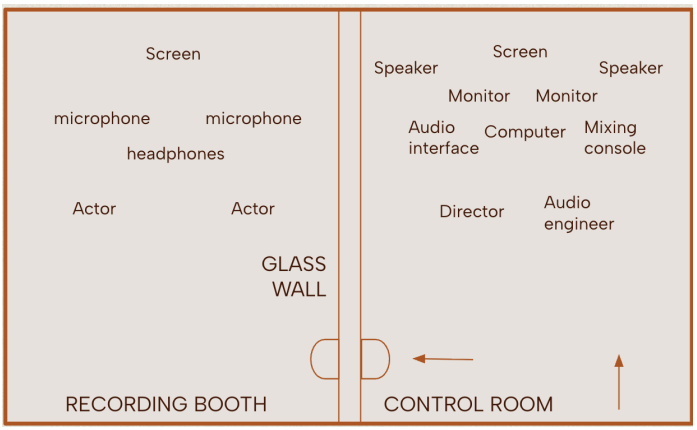
[28] Sebastian Abrahamsson, Filippo Bertoni, Annemarie Mol, and Roque I. Martín, "Living with Omega-3: New Materialism and Enduring Concerns," *Environment and Planning D: Society and Space* 33, no. 1 (2015): 12, 15, https://doi.org/10.1068/d14086p.
[29] *Ibid.*, 10.

Chapter 2 Acoustic Infrastructure

Approaching the studio for the first time, one descends a short flight of stairs and pushes open a heavy, sound-insulated door, the first of several thresholds that mark the transition from the ordinary acoustic world outside to something more carefully controlled within. Following the direction shown in Figure 1, the first space entered is the control room. Directly ahead sits a large desk with a computer and a dual-monitor setup. Beside it are the mixing console and the audio interface, whose backs disappear into a dense web of cables. Studio monitors flank the desk on both sides, their low-distortion speakers designed to reproduce sound with the highest possible fidelity. Mounted on the wall above is a screen that allows staff to view footage during recording projects that require sound-image synchronization. To the left, the wall gives way to a thick pane of glass. What Figure 2 below shows is the place seen through the glass wall. Beyond it lies the recording booth, accessible only by pushing open a second set of heavy, insulated doors. Mirroring the control room's layout, the booth also features a screen for displaying visual content during recording. Positioned in front of it are two tall, large-diaphragm condenser microphones, with several pairs of headphones tangled together and hanging nearby.

For the aspiring voice actor entering this space, the immediate experience is one of orientation without comprehension. The equipment is legible as professional, in its density and specificity, signaling that serious, standardized work happens here. However, its actual operations remain largely unknown to career starters, since their job is more about their voice than about understanding the entire production process. During recording sessions, the actor stands in front of the microphone, wearing headphones, hearing their own voice returned to them through the headphones in a form that sounds subtly but unmistakably different from how it sounds in their head. The sound captured by the microphone is transmitted simultaneously to the actor's ear and to the monitors in the control room via cables threaded

beneath the insulated surfaces. The director and audio engineer, seated outside, communicate with the actor via a small talkback microphone that can be activated at any moment. In these moments, actors and staff exchange cues across the glass. Yet what passes between them is not only instruction and performance. It is also, less visibly, a structural relationship: the actor inside the booth hears what the microphone captures and what the engineer allows through; the engineer outside hears everything, adjusts what the actor cannot, and makes decisions whose technical basis the actor rarely knows. Standing inside this space, it becomes evident that the studio is more than a site of performance. It is an environment that actively organizes what practitioners can sense, know, and do, and, crucially, what remains beyond their perception entirely. This chapter argues that this organizing function is not incidental to the studio's design but constitutive of it.





(Figures 1 and 2: the left is an image I made of the floor plan of the recording studio, and the right is a picture of the recording booth filming through the glass)

To understand this constitutive function, it is necessary to ask what kind of thing the recording studio is. The most immediate answer is that the studio is a room equipped for sound production. Yet, beyond a fruitful description of the studio and what it contains, the

studio in practice often operates according to what it connects, to the practitioners who inhabit it, to the industry standards that govern what it produces, to the institutions that determine who may enter it and on what terms, and to the broader networks of technology, labor, and legitimacy within which its outputs circulate. It is this connective, organizing, and partially invisible function that the concept of infrastructure is designed to capture. Nevertheless, infrastructure, as theorized across a range of disciplines, carries different emphases depending on the scale and object of analysis. Before the recording studio can be understood as an infrastructure or what I term "acoustic infrastructure," it is necessary to trace how the concept has been developed, where its existing formulations fall short for this analysis, and what work a specifically acoustic and institutional account of infrastructure makes possible.

Studio and Infrastructure

It is not fair to theorize the studio as an infrastructure based on a default understanding of large-scale systems that support the operation of state and modernity, as railroad tracks, city plumbing, and electricity do.[30] The studio, equipped with advanced devices, is more of an access-controlled space where professionalism is cultivated through intricate relationships with practitioners than a set of production technologies. It is a relational concept and property in relation to organized practices, namely the voice acting recording session, "not as a thing stripped of use."[31] Similarly, the relational studio as a taken-for-granted background consistently escapes analytical attention, demanding ethnographic scrutiny and reorientation. Star describes infrastructure as "transparent" and "ready-to-hand," which "by definition invisible, part of the background for other kinds of

[30] Brian Larkin, "The Politics and Poetics of Infrastructure," *Annual Review of Anthropology* 42 (2013): 330, https://doi-org.proxy.lib.umich.edu/10.1146/annurev-anthro-092412-155522.; Brian Larkin, *Signal and Noise: Media, Infrastructure, and Urban Culture in Nigeria* (Durham, NC: Duke University Press, 2008),19.
[31] Susan Leigh Star, "The Ethnography of Infrastructure," *American Behavioral Scientist* 43, no. 3 (1999): 380, https://doi.org/10.1177/00027649921955326.

work."[32] The mixing console, the audio interface, and the cables disappearing into the wall are not objects that demand interpretation but are background conditions of a practice whose foreground is entirely occupied by performance. Studying them requires what Star calls confronting an "embedded strangeness," a second-order strangeness of "the forgotten, the background, the frozen in place."[33]

Yet the recording studio complicates Star's account in a way that her framework does not fully anticipate. For Star, embedded strangeness is primarily a methodological problem, in which she calls for ethnographers' sensitivity to recognize the stabilized infrastructural system through long use. In this sense, the invisible characteristics are conceived from a natural consequence of infrastructure's maturation — the more smoothly it operates, the less visible it becomes. What the recording studio reveals, however, is that invisibility can also be actively produced rather than naturally achieved. The students who enter MZ Studio are not long-habituated users for whom the technical environment has gradually faded into the background. They are encountering it for the first time. Their unfamiliarity with the studio's technical operations is not due to overfamiliarity but to a deliberate institutional arrangement. They are taught studio etiquette before signal flow. Hence, the institutional authority of the studio is sustained not only by the full choreography of its material conditions but also by the opacity it maintains toward those who depend on it most.

The attention to unfamiliarity does not suggest a similar interpretation for all practitioners. As Star and Geoffrey C. Bowker demonstrate that the usefulness of infrastructure depends on different populations' judgment,[34] the capacity and power of the studio are similarly imagined and differently experienced by practitioners across different positions, not limited to voice actors, directors, and audio engineers, whom I encounter more

---

[32] Star, 381.
[33] *Ibid.*, 379.
[34] Susan Leigh Star and Geoffrey C. Bowker, "How to Infrastructure," in *Handbook of New Media: Social Shaping and Social Consequences of ICTs*, updated student ed., ed. Leah A. Lievrouw and Sonia Livingstone (London: SAGE Publications, 2010), 230, https://doi.org/10.4135/9781446211304.n13.

often than others. However, in practice, among all working positions, voice actors and their narratives are special and heuristic, given their most frequent relationship with the studio, yet they are the least informed about it. Thus, their structurally estranged interactions with the infrastructure make its presence legible only at moments of breakdown. Drawing from this argument, it reveals the gap between idealized professional proficiency and the contingent, materially embedded conditions under which that proficiency is produced.[35] Based on ethnographic cases from the perspective of a voice-actor trainee, especially how career starters are introduced to the professional working environment for the very first time, I argue that the recording studio, as an acoustic infrastructure, does not merely provide the conditions where professional voice acting takes place, but actively produces the terms by which professionalism is defined and experienced through arrangements that remain largely outside conscious understanding, particularly for voice actors.

Media Infrastructure, Sonic Infrastructure, and Acoustic Infrastructure

Unlike large-scale modernization projects in which infrastructure serves state-building from above, the recording studio, as the infrastructure of voice acting, operates on a more intimate scale. In the Chinese context, its spread was definitely driven by the development of the larger media industry. The recording studio as a site of voice acting is not a recent invention but is historically embedded in the film industry's infrastructure.[36] The legacy continues to shape the genre hierarchy within contemporary training programs, where film dubbing dominates most training practices, even though students are motivated to learn other popular genres, including animation, video games, and Internet radio drama. Therefore, the

[35] Star, 379, 382. In *The Ethnography of Infrastructure*, Star explains nine properties of infrastructure. The following discussion on how students realized the presence of the studio as an infrastructure only when it breaks down addresses two properties: *embeddedness* and *built on an installed base*. As I'll explain, the studio for voice acting inherited habits and standards from other existing industrial developments, such as the film industry and digital video format. I will use a case of how inappropriate video format of practicing pieces reveals how the acoustic infrastructure is an infrastructure and how sound production is not completely isolated from other media forms.
[36] Feng, 83.

studio, as an infrastructure for voice acting today, does not spring up out of nowhere, but is embedded in other structures (such as the film studio), social arrangements, and technologies.[37] As I described before, it is composed of a variety of devices, whether acoustic or not, such as microphones, loudspeakers, headphones, monitors, soundproof walls, and heavy doors, and is a built-up network that relies on other existing technological systems.

In Larkin's discussion on the ontology of infrastructure, he suggests that the duality of infrastructure as both things and the relation between things requires analysis beyond the dominant thinking that "presumes a clear, linear relationship between an underlying system and the phenomenal world."[38] Lisa Parks' framework of media infrastructure rejects the casual relationship, saying infrastructure cannot be reduced to networks alone, but instead foregrounds the material, spatial, and circulatory conditions that make media systems possible.[39] Parks' theorization does not reject networks outright. Instead, it insists that infrastructures must also be understood through their material, physical, and circulatory forms. As she clearly points out, her definition of media infrastructure "refers to the material resources that are arranged and used to distribute audiovisual content," which "extend far beyond the studio and the screen, and include raw materials such as the sun, electricity, land, water, petroleum, chemicals, heavy metals, plastics, and spectrum."[40] Although this research does not address the aforementioned raw materials, Parks' suggestions on material resources still inspire the following discussion. In the recording studio, this means attending not only to what devices are present but to how their spatial arrangement, acoustic treatment, and material configuration collectively shape what kinds of sound production are possible. For instance, patched hardware and software, along with their networked systems, constitute an

---

[37] Star and Bowker, 231.
[38] Larkin, 329.
[39] Lisa Parks, "Stuff You Can Kick: Toward a Theory of Media Infrastructures," in *Between Humanities and the Digital*, ed. Patrik Svensson and David Theo Goldberg (Cambridge, MA: MIT Press, 2015), 356, https://doi-org.proxy.lib.umich.edu/10.7551/mitpress/9465.03.0031.
[40] *Ibid*.

indispensable environment and space for production. Thus, the infrastructural studio is best understood not as a linear chain of technical functions, but as a visible configuration of relations among devices, people, and practices. In addition, operating the recording studio requires practitioners, such as audio engineers and directors, to manage the system and voice actors involved in content creation. The recording studio, with its assemblage of devices and the practitioners who manage them, embodies the convergence of technical systems and social networks, integrating not only artifacts but human habits, norms, and roles.[41] Yet, how voice actors understand the infrastructural studio emerges from a paradoxical position**:** practitioners perceive, as the infrastructure, the studio's existence and authority through what they hear every day, the high-quality recording, instructions from pioneering practitioners, rather than through a clear understanding of what produces it. The unique understanding, then, requires new vocabulary to elaborate on its specificity and the particular context of voice acting. Therefore, I introduce the concept of acoustic infrastructure to capture and highlight the acoustic mediation of infrastructural knowledge, the way sound becomes both the product and the proof of the infrastructure's operation.

Modern sound is the result of technological mediation, especially in the case of the recording studio here. Alongside increasing control over architectural materials, new materials are designed to govern sound behavior, thereby constituting the modern standard of good sound and challenging the relationship between sound, space, and time through acoustic development in modern production.[42] Similarly, the recording studio for voice acting is constituted by layers of acoustic design whose function is not merely to improve sound but to define what professional sound is. The studio is intentionally designed to capture "clean" sound, isolated from its original space and unnecessary background noise, and transmitted

[41] Larkin, 330, 334.; Paul N. Edwards, Geoffrey C. Bowker, Steven J. Jackson, and Robin Williams, "Introduction: An Agenda for Infrastructure Studies," *Journal of the Association for Information Systems* 10, no. 5 (2009): 366, https://doi.org/10.17705/1jais.00200.
[42] Emily Thompson, "Sound, Modernity and History," in *The Sound Studies Reader*, ed. Jonathan Sterne (New York: Routledge, 2012), 118-19.

with minimal interference.[43] Nevertheless, as the following session unfolds, this technical control over the acoustic space extends beyond the physical properties of clean sound. It defines how its users enter, behave, and produce within it. In other words, beyond the technical transformation of the sound signal, the technologized studio also demonstrates mastery over the people involved and their behaviors through its institutional control. This institutional dimension of acoustic control is precisely what existing frameworks for sonic infrastructure in the Chinese context have been unable to theorize.

The concept of sonic infrastructure offers the closest existing framework for understanding how sound organizes social life in the Chinese context, yet it ultimately points toward a different set of questions than what this chapter pursues. In Maoist China, sonic infrastructure operated as a top-down instrument of state power through radios, loudspeakers, and organized listening.[44] As Adam Kielman demonstrates, this top-down political model was subsequently transformed by non-state actors, such as record companies, independent producers, and social networking sites, that shifted its orientation toward individual listeners after the 2000s. His notion of sonic infrastructure illuminates how the space-time relationship and materiality of media systems are embedded in and shaped by dynamics of authoritarian controls and neoliberal developments in contemporary China.[45] In this framework, sonic infrastructure is primarily a circulatory concept that tracks how sound moves through social and political networks, assigns identities and belongings, and reorganizes the relationship between urban and rural subjects. Where sonic infrastructure tracks how sound moves through social and political networks, acoustic infrastructure redirects attention to where sound is made, namely, the bounded, access-controlled space of the studio, and the

[43] Louise Meintjes, "The Recording Studio as Fetish," in *Sound of Africa!: Making Music Zulu in a South African Studio* (Durham: Duke University Press, 2003), 273.; Thompson, 121.
[44] Jie, Li. "Revolutionary Echoes: Radios and Loudspeakers in the Mao Era," *Twentieth-Century China* 45, no. 1 (2020): 27. https://dx.doi.org/10.1353/tcc.2020.0006.
[45] Adam Kielman, "Sonic Infrastructures, Musical Circulation and Listening Practices in a Changing People's Republic of China," *Sound Studies* 4, no. 1 (2018): 24-25, https://doi.org/10.1080/20551940.2018.1499067.

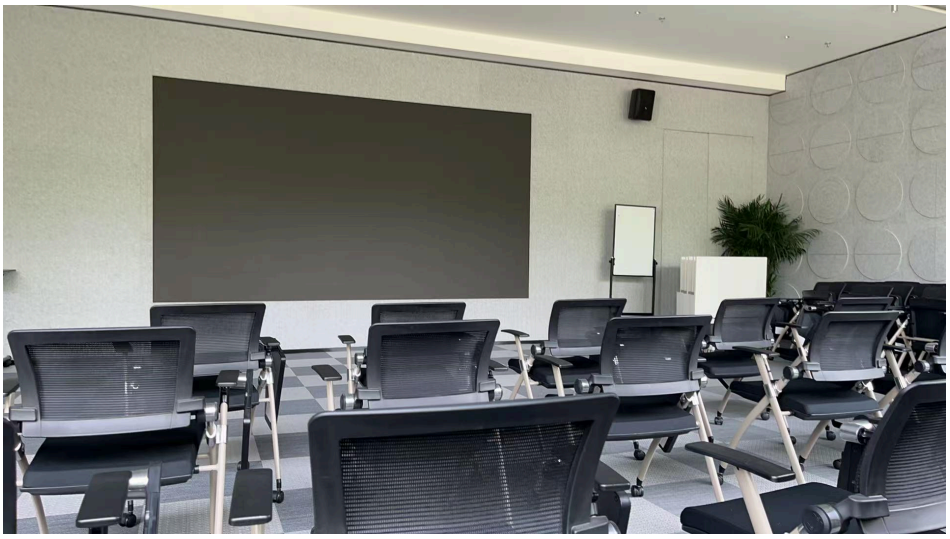
institutional conditions. Therefore, my intervention here emphasizes everyday practices and highlights infrastructure as a knowledge-production system in which positionalities shape what the recording studio is. I define the recording studio as an acoustic infrastructure: a networked system of physical spaces, technologies, institutional arrangements, and social practices through which professional sound is produced, standardized, and recognized as legitimate and through which the practitioners who produce that sound are themselves formed, positioned, and evaluated.

In the following discussion, I will introduce acoustic infrastructure through three dimensions: material/acoustic, institutional, and relational. By bringing in cases of spatial design, studio etiquette, and unexpected accidents, the studio is theorized as an acoustic infrastructure itself and as part of an acoustic infrastructure of audio production, being largely compatible with and connected to other underlying system networks.

The Material/Acoustic Dimension: The Acoustic Space

The very first training session was not held inside the recording booth. The students were guided into a spacious acting classroom with chairs flanking the back, as shown in Figure 3. The floor was cleared for us to make a circle and start practicing speaking and breathing. The 17 of us, coming from different cities across China and meeting in person for the first time, were nervous about raising our voices, and, of course, the room's reverberation only amplified our hesitation. Passing by everyone and back to the center of the circle we made, the instructor brought up his concern: "We'll have a final showcase here on our last day. By that time, remember to speak as loudly as you can. Otherwise, your voice will not be heard clearly even with stand-front microphones." Later, when our training sessions were moved to the recording booth, the voices transmitted through the microphones and speakers remained loud and clear. Compared to the meeting classroom, which requires no acoustic

design for daily use, the recording studio is a space of deliberate acoustic intervention — every surface, every material, every threshold engineered to produce a particular kind of sound.



(Figure 3: the large classroom where we started the program and held the final showcase. When we were holding sessions, the chairs were pushed to the back, leaving a vacant floor for us to make a circle)

As the previous floor plan shows, the control room is the entrance point of the recording studio. Over the thick glass wall lies the recording booth, a typical room-within-a-room construction in acoustic design.[46] Beyond the glass is the isolated production space, while the other side, as seen from the filming angle in Figure 4, is arranged with devices. The double-layered transparent glass wall, as shown in Figure 4, allows the actors and director to see each other without sound penetrating through it. Similarly, access to the inner room is through two thick soundproof doors. By creating a heavy soundproof gate, the inner recording booth is isolated from the outside space, keeping unwanted sound out and preventing sound from passing between rooms.[47] Beyond isolation, the booth's interior surfaces are treated with a coordinated system of acoustic materials. For example, as shown in Figure 5, bass traps placed in the corners absorb low-frequency buildup, absorption pads

[46] Ben Harris, "Introduction to Acoustics," in *Home Studio Setup: Everything You Need to Know from Equipment to Acoustics* (New York: Routledge, 2009), 13, https://doi.org/10.1016/B978-0-240-81134-5.00001-2.
[47] *Ibid.*, 12.

on the walls eliminate unwanted reflections across the mid-to-high-frequency range, and diffusion panels on the walls and ceilings scatter acoustic energy evenly to prevent it from concentrating at the microphone position. Together, these treatments address the reverb and room-mode problems that would otherwise interrupt the recorded sound.[48] The cumulative effect of these acoustic treatments is not simply to improve sound quality but to produce a fundamentally different sonic ontology, in which sound is detached from the spatial signatures of the environment that produced it.



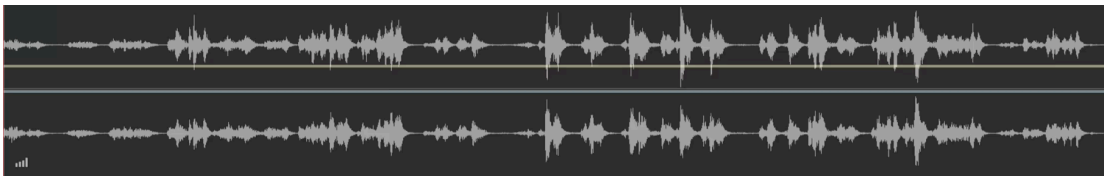

(Figures 4 and 5: the left image shows the control room, where we can see through the recording booth through the glass wall; the right image shows the bass traps in the corner and diffusion panels on the ceiling.)

During recording, the actor is placed in a space that is physically free of unwanted sound, or as close to that condition as acoustic engineering can achieve. The sound collected on the track, in a technical sense, is dead, stripped of its original reflections. This clean, dry recording primarily serves as the industry default. Yet its acoustic distinctiveness is difficult to perceive and only becomes legible through comparison with work produced in other acoustic environments. The recording studio's acoustic character differs fundamentally from the spaces practitioners encounter in everyday life, and it is precisely this difference that causes studio sound to be fetishized as something special — something professional.[49] This fetishization, however, does not mean that the studio's equipment can produce this effect

[48] Harris, 6-8.
[49] Meintjes, 272.

independently. A comparison between studio-recorded soundtracks, as shown in Figure 6, and classroom performances, as shown in Figure 7, makes this clear, since in both settings a microphone serves as the primary interface between the performer's voice and the recording. Yet the sounds produced in these two settings are acoustically distinct, in both auditory perception and acoustic measurement, in ways that cannot be attributed solely to performance but are strongly negotiated by the production environment.





(Figures 6 and 7: the above track is recorded in the recording booth with final comping, see Audio Example 1. The one below is the recording extracted from videos taken during the final showcase in the classroom, see Audio Example 2)

Without acoustic absorption materials, the classroom setting preserves the liveness of the performance and the spatial signatures of the room, its reflections, its reverberation, and its ambient noise. What changes between the two settings is not the performer, not the microphone, but the acoustic environment itself. The presence of a microphone in both settings demonstrates that hardware alone does not guarantee a professional-sounding result. Rather, professionalism is relationally produced, only when the body, the voice, the microphone, and the studio's networked acoustic devices operate together, that the sound is recognized by default as professional. The studio is not the container of professional practice, but one of its constitutive conditions.

The Institutional Dimension: For the First Time

On the third day of the program, we were officially introduced to the recording studio. The instructor for the first half of the session was a program alum who graduated two years ago and has already become a rising star in the industry, with her own voice studio. As we crowded into the control room, feeling curious about the well-equipped environment, she reminded us of something else beyond the spectacular technologies. According to my class notes, she mentioned a list of "dos" and "don'ts":

1. Remember to practice your fundamentals before every recording session, such as vocal techniques, articulatory agility, and breath control.

2. Do not wear clothes or jewelry that rattles or makes a rubbing sound.

3. Do not eat in the studio, and if you want to bring your own water, make sure the bottle has a lid.

4. Remember to practice and get familiar with the piece you're going to record – the working time is precious.

5. Do not touch the devices without permission; if you feel uncomfortable with the microphone's positioning, ask the audio engineer for help. Do not move it by yourself.

It seems fair that, as voice actors, there is no more need for us to learn how the technical system actually works than to memorize the studio etiquette. In other words, the main requirement is not to interrupt the recording session with other unwanted sounds, except for the vocal performance; just close the door, put on the headphones, stand in front of the microphone, keep a certain distance from it, and start.

What this list enacts is a substitution: behavioral norms replace technical knowledge as the primary content of professional initiation. The underlying reasons for the studio arrangement, such as how and where to position the microphone, remain unknown to voice actors who work with them and are most frequently ignored. The instructed interactions

constrain practitioners' access to the material conditions of sound and the technical processes through which their voices are captured and transformed into digital data for circulation. The studio's limited exposure in everyday interactions reveals the infrastructure only partially to some of its users. The infrastructural nature of the studio is evident not only in its invisibility but also in its visibility to specific groups. As the director and audio engineer sit in the control room and manage the soundtracks while voice actors stay blinded, specific devices and the mechanisms for collecting, managing, and even fixing problems with sound recording are only "visible" and "transparent" to specialists. As a result, the studio is understood, overlooked, and sometimes simplified as a professional space through the very limits placed on what students are allowed to see and know. The studio's institutional authority is sustained not only by its acoustic materials but also by the institutionally managed boundary between what practitioners are expected to know and what they are not.

After the instructor introduced the expected behaviors in the studios, some of us were already jumping at the chance to record. While the first group rushed into the recording booth, the instructor further explained what to do for the rest of us waiting outside.

"As early career starters, you probably won't have the chance to perform main roles. Most of the time, you will be assigned to a supporting role. When you sit in the control room and wait for your session, do not miss the opportunity to learn from the seniors' recording inside. Auditing outside, you will learn what recordings sound like outside compared to what you would hear through headphones inside. Remember to pay attention to how the director refines their performance. Once you learn it long enough, you will get what a professional standard is."

In other words, the standard of professionalism is not only transmitted through technical explanation but through the spatial practice of listening from the right position. Since soundproof doors and thick glass walls separate the control room and the recording booth, the director will instruct through an on-desk mic, transmitting his voice to the actors' headphones. Subsequently, the actors' voices inside will be carried over patched cables and

played through the speakers outside. The soundproof architecture acoustically enhances the inside-outside distinction, while the voice acting production requires a deliberate breakthrough to allow the inside sound to flow to the outside. In this sense, the inside-outside "information gap" initiated by the gatekeeping soundproof materials is eliminated even when they signify the closedness of a supposedly symbolic door between two spaces.[50] However, the audit action for career starters to gain experience leaves the condition in an oscillating state. The recording studio is open to practitioners, not limited to voice actors, directors, and audio engineers, as I previously mentioned, in the voice acting industry. Put simply, beyond the control room door, the space is designated for professional production. Yet, when career starters audit the recording session, they intrude on the professional sphere with less experience than the seniors recording inside. In this sense, the audit presents a symbolic boundary between the professional and the rest, while keeping both together to present a unity of larger distinction against non-voice-acting practitioners. Yet the audit is not intended to reinforce the distinction but rather to envision a possible progression – to cultivate career starters into future professionals.

Sitting silently behind the director and the audio engineers, students who are not on call may listen to others' performances. The professional skills do not come from self-discovery or self-practice. In contrast, it emerges from a continuing process of learning that "involves acts of making, improvisation and adjusting to change" and is "a profoundly embodied experience, not just located in cognition."[51] As Thomas Porcello has shown in music recording contexts, sound engineering professionals "require(s) competence both in knowing the linguistic resources and in being aware of what constitutes their appropriate and inappropriate use within both the professional community and the particular situated

[50] Bernhard Siegert, "Doors: On the Materiality of the Symbolic," trans. John Durham Peters, *Grey Room*, no. 47 (2012): 15, https://doi.org/10.1162/GREY_a_00067.
[51] Deborah Lupton, "How Do Data Come to Matter? Living and Becoming with Personal Data," *Big Data & Society* 5, no. 2 (2018): 4, https://doi.org/10.1177/2053951718786314.

context."[52] However, in the Chinese voice acting context, voice actors are not initiated into the language of acoustic infrastructure at all. Without being instructed in the technical language of studio acoustics, signal processing, or microphone selection, their professional identity is constituted not only by their own vocal performances but also by embodied presence and the lived experience of listening to others, rather than by technical knowledge of the environment that produces the sound. This mode of learning is not incidental but institutionally endorsed. As one of the instructors explains, "career starters should cherish the opportunity to audit sessions of professional voice actors, during which you will learn about how they perform and how directors instruct them to adjust to different requirements quickly."[53] By adopting the director's position in the control room, students observe and listen to real-time performance, thus internalizing the industry's aesthetic standards. The spatial arrangement and particular power dynamics afforded by the auditing opportunities simultaneously create and dissolve the standard of being inside as a professional, and of being outside as not. By choreographing institutional power, even if the actors in the recording booth are not experienced seniors, the studio still encodes professionalism by extending the temporality of becoming it through the very circulation of sound and authority between the two sides.

The Relational Dimension: When the Infrastructure Breaks Down

In the previous chapter, I mentioned that students usually select and submit practice pieces to the instructor before the day of recording if the instructor has not assigned pieces. The program aims to cover mainstream genres of voice acting in the market, so that students are prepared for the competitive industry. Spanning topics across films, modern TV dramas,

[52] Thomas Porcello, "Speaking of Sound: Language and the Professionalization of Sound-Recording Engineers," *Social Studies of Science* 34, no. 5 (2004): 740, https://doi.org/10.1177/0306312704047328.
[53] "professional voice actors" here is a direct translation of the instructor's words. The professionals here are experienced senior voice actors in the industry, as the counterpart of us, career starters with little experience working in the studio.

costume dramas, cartoons, anime, video games, and Internet radio dramas, audiovisual voice acting constitutes the majority of the training. That means, students would either download popular clips online for convenience or make screen recordings of their favorite works and submit them to the instructor. The next day, when the recording session starts, the instructor and the audio engineer will click on each clip, ensuring that they work on both screens in the control room and the recording booth. Then, they will ask the "owner" of the clip to start practicing. The take-turn practice works smoothly until something goes wrong with the video clips: one of them seems incompatible with the monitors. After several attempts to open the file, the instructor asked the next student to record in advance while the staff and the audio engineer checked and fixed the problematic clip.

This unexpected occurrence reveals how the studio is an acoustic infrastructure not only for its acoustics but also for its role as a node in the network embedded within the larger infrastructure of digital media production. Equipped with monitors, an audio interface, speakers, various microphones, headphones, and tangled cables behind the desk, the studio is much more than the devices I can see or name. The invisible standard embedded in systems, including file formats, codec requirements, playback compatibility, and other unnoticed requirements, is uncovered when backstage operations accidentally leave traces and demand improvised adjustments.[54] The video file format becomes important and perceptible when it cannot be processed, further reinforcing its invisible yet ordinary role in supporting the production process.[55] The studio is not an enclosed, self-sufficient network or space. Instead, it naturally and stealthily connects to and interacts with students' personal devices, which they use to download or edit practice clips, the file format the monitor can process, and technical standards for industry-wide use. In other words, the studio is by no means an isolated production environment, and its boundary naturally and invisibly extends outward.

[54] Star, 385.
[55] Sterne, 832.

As the recording session encounters an unexpected moment that reveals the underlying secrets of how the digital signal functions, so too is the role of the audio engineer reintroduced. In the previous paragraphs, I consistently mentioned the audio engineer as an equally important role as the director plays in the training program. Nevertheless, as trained voice actors, our students' contact with the audio engineer is limited to the moment he shares the finalized recording files in the group chat at the end of each session. His efforts are largely appreciated, but at the same time, are easily forgotten. Until the session is interrupted, the audio engineer is responsible for fixing the problem, further suggesting his role in maintaining the normal operation, and students or even practitioners in other positions do not need to know. The partial visibility/invisibility of the studio here does not warrant criticism of its black box status. Rather, from the audio engineer's perspective, the acoustic infrastructure here operates inherently differently from what others know and see. While voice actors and directors focus on the artistry of performance, the audio engineer attends to the technical standards that ensure audio files can circulate within the industry.[56] Hence, the operation of acoustic infrastructure involves not only acoustic conditions but also a broader network of technologies and institutions, in which the proper operation of all nodes of this network produces sound. The acoustic infrastructure presents itself to different groups: its passive users, directors, managers, as well as its backstage fixers. What the breakdown reveals, then, is that the acoustic infrastructure is not a single unified system but a set of overlapping and partially visible networks, each legible to some practitioners and opaque to others.

Coda

When did I finally find out the invisible digital format? It was when I had graduated from the program, and students were asked to collaborate on an Internet radio drama as the

[56] Shannon Mattern, "Maintenance and Care," Places Journal, November 2018, 2, https://doi.org/10.22269/181120.; Meintjes, 275.

final showcase. Other students were excited to rush back to Beijing and return to the familiar studio for the professional moment. While studying in the US, I was not able to join, nor did I have recording devices in my apartment. After checking with the instructor, I suggested I could book a recording booth on campus and record my part with the "professional" quality these normally inaccessible devices deliver. Then I had a hard time managing everything by myself: opening Pro Tools, connecting the soundtrack with the in-booth microphone, ensuring I could hear myself in the headphones, and trying not to make noise; otherwise, I didn't know how to find and reduce them in tracks. Although I consider myself a sound studies scholar, there are still procedures my theoretical training does not prepare me for. And when I finally began performing, the strangeness of recording in isolation, without a director's talkback voice, without other actors breathing in the same shared temporality, made the performance feel unfamiliar, almost disembodied. I then wondered what this would sound like when editing those recordings together, far away from Beijing. This frustrating moment crystallized a realization running through this paper. The recording studio is not merely a site where sound is captured; it is the infrastructural condition that organizes how professionalism is sensed, enacted, and believed.

Taken together, the three dimensions of acoustic infrastructure examined in this chapter, the material-acoustic, the institutional, and the relational, are not separate features of the recording studio but interlocking conditions that together produce a professional environment that remains, for most voice actors, only partially visible. The studio's soundproofed walls define what sounds are possible; its etiquette and spatial arrangements define what practitioners are expected to know; and its embeddedness in broader networks of digital standards defines what counts as a deliverable professional product. What connects all three is the same underlying dynamic: the conditions that make professional voice acting possible are precisely the conditions that practitioners are least equipped to see.

Yet this chapter has not yet addressed what happens within this environment — not to the acoustic materials, but to the body that enters it. If the acoustic infrastructure sets the terms of professional production, the question that follows is how practitioners come to inhabit those underlying standards: how they adjust their voices, their posture, their breath, and their sense of self to fit a set of technical conditions whose full logic remains opaque to them. This hanging question will be discussed in the following chapter, the making of the technovocalic body.

Chapter 3 The Technovocalic Body

"Damn it! How can you do this to the stage? How can you do this to the fans who love you? Tell me!"

"Stop. Selina, this part is too 'watery'.[57] Try to be more emotional, be angry. And don't forget to give me 'layers' of expressions. Start over!"

Mr. Zhang, the instructor of the training program, interrupted my performance. At the time, I was dubbing the role of a traditional Chinese drama master interrogating her student. The student character, usually highly focused, had been distracted on stage that day due to her longing for an ideal romance. Hearing Mr. Zhang's advice, I tried hard to organize my thoughts, wondering what "watery" means and why. Feeling sorry, but still, I held my headphones in place and asked with sincerity:

"Mr. Zhang, could you tell me what you mean by that and what layers?"

"Well, for now, it just sounds like *you* are talking to us, not *the character* as a master. It's too natural and lifelike. For sure, being natural in performance is a good thing, but you still have to have some 'gestures'.[58] See the clip and see when the actress started to fly in a rage. Just follow the way her body moves and facial expressions. Again."

I memorized when the actress paused, opened her mouth, trembled her lips, and raised her voice. I clenched my fists, controlled my breath, and tensed my muscles, if I had any, when I performed, feeling like I had made some adjustments and made the lines more furious.

---

[57] The original context here is the Chinese character "水 (*shui*)", which means water when it is a noun. Here, the instructor was using "水" as an adjective, which I translated it as "watery", meaning too flat and without undulations.

[58] The term "gesture" here does not mean the physical movement of the body, but more like a tone. It is translated from the Chinese character "范儿/*fan er*", which is usually translated as "style". However, I pick "gesture" as the corresponding one since the tone Mr. Zhang requires is the feeling of a voice with ornamented skills from the systematically trained body, just like how gestures serve as a supporting visualization for emotional delivery.

"Ok. Let's listen to the playback…… I still didn't get your changes. It's just the same as your previous take. Anyway, time's up. Try to practice more after this session. Next group!"

After thanking Mr. Zhang and the audio engineer who helped with the recording, I took off the headphones and sat outside the studio, holding the script and mulling over the notes. The big question almost jumped out from the lines and my confusing facial expression: why didn't Mr. Zhang get my adjustment while I've "heard" it in my headphones?

Extending the previous chapter's discussion of the studio as an acoustic infrastructure, my confusion can be partially addressed through an acoustic lens. As I noted earlier, during a recording session, students stand inside the recording booth in front of the microphone, wearing a pair of headphones. At the same time, their ears are simultaneously fed the video's vocal-removed audio track and their own mediated voice as they speak. At the same time, the microphone, routed to the monitors in the external control room out of sight, transmits the performance to the director and audio engineer. The sound, stripped as much as possible of the recording booth's acoustic properties, is projected through the speakers into another space: the control room. So the director and the audio engineer hear the sound as it merges with the spatial acoustics of the control room, rather than the isolated, decontextualized feed the voice actor hears, the same sound, but has lost its detachable echoes.[59] Conversely, students inside hear their own voice differently from the familiar resonance of bone conduction. This is not to say that the recording equipment has fundamentally altered the track's sonic properties. Rather, compared to the wall-mounted monitors, the headphones, by isolating the left and right ears, not only preserve the spaceless, clean sound but also

[59] Jonathan Sterne, "Space within Space: Artificial Reverb and the Detachable Echo," *Grey Room*, no. 60 (Summer 2015): 114, 116, https://www.jstor.org/stable/43832233.

acoustically detach the student from the physical space they inhabit, creating an experience entirely distinct from hearing one's own voice in an unmediated room or through speakers.[60]

The above is a discussion of how the auditory experiences of both parties are theoretically organized. In reality, human ears do not necessarily process these nuanced spatial acoustics so consciously. Even if they did, acoustic properties alone do not determine the quality of a voice acting performance. As shown in the case above, Mr. Zhang suggested removing the "me" part of my voice to perform the character, whereas I was confused because I thought my body had adjusted my voice, though the result was unsuccessful. Ultimately, this is a question of embodied performance rather than pure hardware. Grounded in the perspective of career starters, this chapter explores how voice actors physically perform within this acoustic infrastructure, examining how practitioners, whether fully aware of or largely estranged from the technical environment, adapt to achieve a professional sound. To address this question, this chapter develops the concept of the technovocalic body, the embodied vocal practice that emerges from and is shaped by the practitioner's interaction with recording technologies. By developing the concept in dialogue with existing scholarship on voice and body-technology relations, I argue that the professional voice in Chinese voice acting is not a natural expression of interior feeling but a materially and technically produced artifact, shaped by conditions that, for most practitioners, remain only partially visible.

The Voice Behind the Screen

In film studies, sound has frequently been theorized through its phenomenal effects on the audience, often at the expense of its material production. The most prominent example of this paradigm is Michel Chion's influential concept of acousmatic sound. The acousmatic

[60] Paul Roquet, "Acoustics of the One Person Space: Headphone Listening, Detachable Ambience, and the Binaural Prehistory of VR," *Sound Studies* 7, no. 1 (2021): 43-44, https://doi.org/10.1080/20551940.2020.1750270.

sound, he defines, refers to sounds "one hears without seeing their originating cause."[61] He further defines acousmatic sound with a comparison to offscreen, arguing that "a sound coming from offscreen is always acousmatic, but the opposite is not necessarily true."[62] Offscreen sound refers to sound originating outside the cinematic frame, whereas acousmatic sound places greater emphasis on situations, "either a sound is visualized first and subsequently acousmatized, or it is acousmatic to start with and is visualized only afterward," in which the sound lacks a corresponding visible source on screen.[63] However, this discrepancy is not a flaw. It is a common cinematic technique. For instance, in the example of *M* mentioned by Chion, withholding the image of the child murderer while allowing his voice to linger continuously creates a sense of mystery.[64] Even without revealing the murderer's visual presence beforehand, his voice, or the eventual synthesis of sound and image, realizes the voice-body synchronization. In contrast, voice acting, as a kind of sound that is neither within the frame of a movie nor visible to the audience as its source, namely the voice actor, is essential for convincing the audience that it is synchronized with the image, leading them to believe that the sound emanates from the image on screen. In other words, voice actors are not the featured visual subjects. Their task is to match their voices to the on-screen actors to create the illusion that the voice actor's sound emanates directly from the screen actor, thus aligning with Chion's further explanation of offscreen sound being deceptive and "is entirely a product of the combination of the visual and aural."[65]

Chion's analysis of sound is largely rooted in the aesthetics of how audiences appreciate film texts and how these materials initiate their psychological interpretation. But to realize this aesthetic, it is not simply a matter of how to (or not) feature screen actors

---

[61] Michel Chion, *Audio-Vision: Sound on Screen*, trans. Claudia Gorbman (New York: Columbia University Press, 2019), 72.
[62] *Ibid.*
[63] *Ibid.*
[64] *Ibid.*, 73.
[65] *Ibid.*, 81.

audiovisually. That is to say, audiovisual synchronization is not merely a simultaneous occurrence, and the affective charge of the voice does not simply emanate from the natural voice itself. Instead, the illusion of naturality is produced through concrete, conditional labor: behind the screen, voice actors detach their vocal performance from their visible bodies. And, they have to go through the specific technical and institutional conditions of the recording studio. As Nozawa suggests, by acknowledging the underlying intervention of technology, voice acting should be considered as a sociotechnical hybrid, in which the voice actor's performance is inseparable from the technological systems that capture, process, and deliver it.[66] Specifically, this form of labor is tightly integrated with the acoustic infrastructure I've defined before, delimiting professional sound through the contingent interactions evident in their mundane practices. As human bodies are simultaneously the instrument, the performer, and the site where technical and expressive requirements must be reconciled, the sound they produce should satisfy both the director's aesthetic judgment and the recording system's technical parameters. Voice acting is, in this sense, not only a form of professional labor but an unusually direct body-technology interface.

Technovocalic Body/Technovocality

When Nozawa introduces the concept of the sociotechnical hybrid, he places voice acting alongside ventriloquism, arguing that both involve techniques and imagination for shaping a character.[67] A ventriloquist produces a voice and attributes it to a puppet, effectively enveloping the inanimate object in a vocalic body. Similarly, voice acting lends the voice actor's voice to a character, either creating an entirely new artificial body or merging with an existing visual presence to achieve complete audio-visual synchronization. Ventriloquism, in this sense, allows inanimate matter to become a vocalic body. Building on

[66] Nozawa, 173.
[67] *Ibid*.

this concept, voice acting resembles a form of digital ventriloquism: under the affordances and constraints of media technologies, voice acting reconfigures the audio-visual relationship between the voice actor and the "body" they work with (whether that body is the image of a screen actor or a virtual character born from a text) and in doing so produces a technovocalic body.[68]

Yet to place voice acting alongside digital ventriloquism is reductive and oversimplifies the complexity of the voice acting process. In existing discussions of digital ventriloquism, the technovocalic body is always a result of how media technologies redistribute voice. In her discussion of deepfake videos, Jaimie Baron proposes the idea of archival ventriloquism. For instance, she explains with the practice of reconfiguring existing video and audio material related to Obama, including the appropriation of his political symbolism, to make him appear to say things he never actively expressed.[69] In this way, his voice becomes detached from its original context and reattributed to a body, an imagined Obama, that never produced it. Subsequently, digital ventriloquism grants the original speaking body a new, unauthorized performance, which is a reconfiguration of the ownership of voice and the control over identity. Another category of digital ventriloquism emerges from discussions of artificial intelligence, particularly regarding AI news anchors and voice assistants. By decomposing human voices, such as those of voice assistants' voice actors or news anchors, into phoneme data, algorithmic ventriloquism appropriates pre-recorded, datafied human voices and reassembles them to construct a digital body.[70] As with the Obama example, algorithmic ventriloquism has been criticized for transferring power over voice and control of it to AI technology companies, thereby weakening and even exploiting the original

[68] Baron, 200.
[69] *Ibid*, 199, 201.
[70] Ido Ramati, "Algorithmic Ventriloquism: The Contested State of Voice in AI Speech Generators," *Social Media + Society* 10, no. 1 (2024): 3, https://doi.org/10.1177/20563051231224401.; Hadar Levy-Landesberg and Xuenan Cao, "Anchoring Voices: The News Anchor's Voice in China from Television to AI," *Media, Culture & Society* 47, no. 2 (2025): 245, https://doi.org/10.1177/01634437241270937.

speaker's agency over their own voice.[71] In both forms of digital ventriloquism, the process of producing a technovocalic body involves an absolute division between the active and the passive: either one loses ownership of one's voice, or one seizes control of a voice and asserts the subjectivity of a body.[72]

Unlike the two forms of digital ventriloquism discussed above, voice acting does not result in the practitioner losing control over their body or voice. On the contrary, it is precisely in the action of voice acting that practitioners must actively reconfigure their own bodies. The body, in other words, is itself a technovocalic body, one that is intervened upon by the recording studio in real time, in the situated moment when sound is being produced. Moreover, the recording studio is the acoustic infrastructure that mediates their interaction while simultaneously rendering it invisible: voice actors attend to their performance, remaining unaware of the studio's shaping force, the bodily labor it affords, and the institutional demands it imposes. As I mentioned in the Introduction chapter, students from art academies argued that school training cannot meet their expectations for voice acting, claiming that their teachers are less attentive to emotional performance. The underlying message here is that students consider the ability to portray characters' emotions well as the priority in voice acting. Still, they rarely mention the specific efforts they put into mobilizing their bodies in the studio. Rather than treating voice acting as either stylized artifice or natural expression, as Amanda Weidman had criticized the linguistic traditions in English, which consider material aspects of voice as secondary, voice is better understood as an already material event produced through the interaction of the body with its surrounding environment, shaped by specific conditions, including technological mediation.[73] A professional voice is not something that voice actors are born with or already possess.

---

[71] Ramati, 6.
[72] Baron, 201.
[73] Nina Sun Eidsheim, *Sensing Sound: Singing and Listening as Vibrational Practice* (Durham: Duke University Press, 2015), 2.; Amanda Weidman, "Voice," in *Keywords in Sound*, ed. David Novak and Matt Sakakeeny (Durham: Duke University Press, 2015), 233.

Instead, it is something that must be learned, trained, and produced every moment the practitioner steps into the recording booth. Mr. Zhang's instruction to remove the "me" part of the voice and perform the character is a demand to consider the importance of the body in vocal production. In the following discussion, I will analyze my theorization of the technovocalic body with two moments: the conscious disciplining of the body and the adaptation of performance habits to technological constraints, which I term the body that is full of memory, and the unconscious construction of a professional sound through processes that practitioners neither control nor fully perceive, which is a digital Frankenstein of sound. It is this corporeal mediation, how the body learns, adjusts, and sometimes fails to adjust, within the studio's technical conditions, that develops the framework of the technovocalic body in a context beyond its existing applications.

The Body Full of Memory

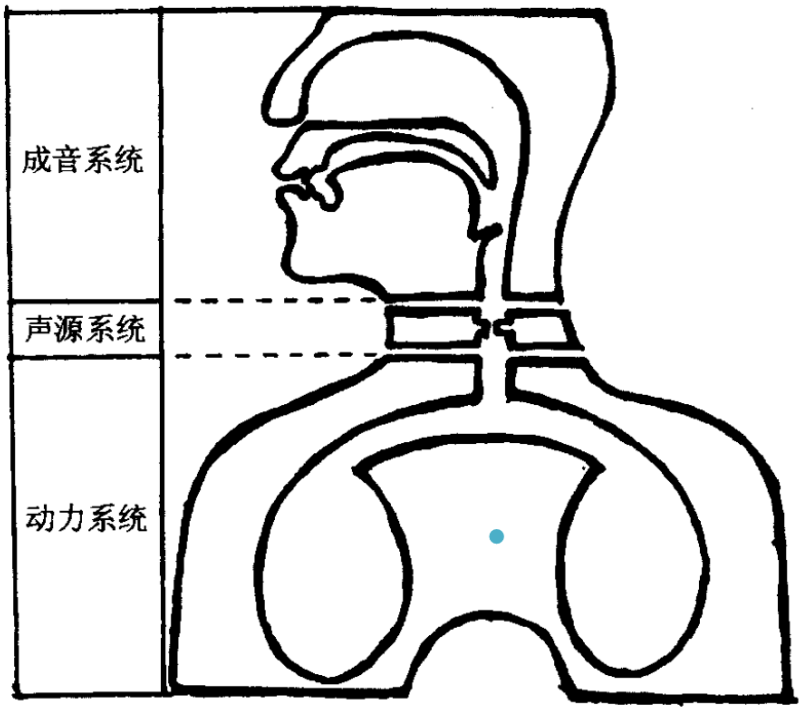
Due to the program's short length and the fact that the other students already had a foundation in acting, we didn't spend much time on vocalization or acting fundamentals. From the third day onwards, our training happened in the studio. As Mr. Zhang said, one of the great things about this program is the experience of learning in a professional environment, while the basic skills can be practiced after class and need to be maintained at all times. For this reason, Mr. Zhang recommends the book *The Skill of Actor Speech Arts*, the official textbook of the Central Academy of Drama and a required reading at many art schools across the country. The book has four main sections: basic language training, including breathing, vocalization, and diction; stage speech performance, such as how to make lines different from just reading them through emphasis, pauses, and intonation; external techniques, which focuses on how to express emotions in different situations and make performances more realistic; and the art of the microphone, which is also known as

dubbing. As shown in the book's introduction, it is a textbook for stage theater performances, and voice acting in a recording studio is only briefly mentioned. What's more, in this book, the difference between voice acting and stage acting is simply whether a microphone is used and how often. This explains why some students have told me that they don't think school training is what they think voice acting is. As I mentioned earlier, voice acting is not a separate, systematically categorized discipline in art schools. In many programs, voice acting is just an elective. According to students, most of the voice acting they've been exposed to in school has been in the service of film and television drama ADR, or documentaries. These types of productions typically require lip syncing or standardized pronunciation. In contrast, students were more interested in less traditional genres, such as video games and Internet radio dramas.

Such works portray non-existent fictional characters without a corporeal body, rather than those performed by screen actors. Then, voice actors' jobs are to become a nobody, which, in Nozawa's research, his interlocutor, a noble voice actor in Japan, describes as spirit mediumship, as if she lends her vocal organ, mouth, and body to the character to allow the character to speak.[74] Although the students aspired to non-traditional voice acting, the lack of a specialized voice acting discipline forced them to borrow, or even patch together, training from other related disciplines. After all, clear lines presuppose clear pronunciation. So, in addition to theater acting textbooks, students also need to train in pronunciation according to broadcasting textbooks. Zhang Song, the first dean of the School of Broadcasting at the Beijing Broadcasting Institute (now Communication University of China), laid the theoretical foundations for the scientificization of the broadcasting art with his 1993 publication, *China Broadcasting Science*, which aimed to standardize the pronunciation of the anchors and to enforce an idealized, and arguably artificially set intonation.[75] Through pictures of anatomical

[74] Nozawa 190.
[75] Levy-Landesberg and Cao, 239.

structures and technical terminology, such as non-anthropomorphic organ diagrams and the categorization of the organs involved in vocalization into the dynamical system (the organs and muscles that provide airflow for vocalization, such as the lungs and the diaphragm), the source system (the larynx and the vocal folds), and the phonatory system (the organs that are involved in resonance, such as the nasal passages and the oral cavity) shown in Figure 8, the science of podcasting establishes an instrumental view of the body.[76]



发音器官示意图

(Figure 8: the diagram in Practical Broadcasting Tutorial vol.1 shows the dynamical system (the bottom section), the source system (the middle section), and the phonatory system (the upper section)

Later, in 2002, the Communication University of China (CUC) published a series of textbooks on practical broadcasting that added several examples to Zhang's training theories, such as word combinations and tongue-twisting segments. Repeated practice of these examples helps improve control of muscles and organs, and thus achieves "automation" (e.g.,

[76] Levy-Landesberg and Cao, 237.; Wu Hongyi 吴弘毅, *Shiyong Boyin Jiaocheng 1* 实用播音教程1 Practical Broadcasting Tutorial, vol. 1 (Beijing: Communication University of China Press, 2002), 254-55.

the book refers to the flexibility of breathing as automated breathing).[77] Thus, voice acting, akin to broadcasting, is not achieved by simply breathing and reading a script, but by emphasizing the materiality of the human voice and language, the necessary involvement of bodily knowledge and training. What is particularly obvious in the example I presented earlier is that Mr. Zhang's comment that my voice was too "watery" or natural was actually a reference to my lack of body control. Since I was the only student who did not major in acting or broadcasting and did not intend to make a career out of voice acting, I had not studied voice training before joining the program. Even though I began practicing these examples during the program, the brief exposure was not enough to familiarize me with my body or effectively control it. And this internal body skill cannot be acquired through instruction or by working in the studio alone. As Weidman argues, voice is a material event requiring an internal choreography within the body.[78] What I lacked was not emotion but the trained capacity to translate that emotion into a sound that met the technical and aesthetic parameters of the studio environment. The body that enters the recording studio is not a neutral instrument. Instead, it arrives already shaped, or not yet shaped, by years of prior training. At the same time, it is this prior shaping, or its absence, that determines what the practitioner can and cannot do when the professional recording session begins.

The internal choreography of the body, guided by broadcasting, and the external choreography required by theater performance, such as interaction with the stage space, have together shaped the theoretical foundation of the current profession of voice acting. Yet when these two forms of training encounter the specific technical demands of the recording studio, their limits become visible, and sometimes, they actively work against each other. This clash became evident in the performance of one of the students, a stage-acting major at one of the top art academies. During an exercise requiring a highly emotional, angry performance, she

[77] Wu, 265.
[78] Weidman, 235

had to do multiple takes. Mr. Zhang told her that her voice was too strong and too loud — it was peaking the microphone, and her breathing was far too obvious. However, when she tried to perform with less physical force, the instructor said her voice sounded too weak, indicating she wasn't fully committing to the emotion. After a lengthy discussion, they concluded that the issue stemmed from her stage acting habits. Stage acting requires a voice that can reach the audience sitting in the back row, even without relying entirely on a microphone. Her body was full of the muscle memory of being on stage, which initially betrayed her in the acoustic isolation of the studio. When she first recorded in there, she used this exact same on-stage technique. But voice acting requires actors to maintain a specific distance from the microphone — ensuring optimal sound capture while preventing the volume from overloading the microphone. This student then made a simple adjustment: she took a few steps backward to increase her physical distance from the microphone. Then she used her habitual, forceful vocal projection to express the anger. This way, her voice retained its emotional power while staying within the microphone's acceptable frequency range, without being so far away that it couldn't be recorded.

This case illustrates something important about the technovocalic body. The student entered the recording studio carrying a set of habits trained for a completely different acoustic environment. Those habits — the projection, the force, the physical commitment — were not wrong in themselves. What the studio required was not a different emotion, but a different body: one that had learned to negotiate between expressive demands and technical parameters. As Sterne argues, MP3 as a technological artifact, "is an item that 'works for' and is 'worked on' by a host of people, ideologies, technologies, and other social and material elements."[79]

[79] Sterne, 826.

Sound reproduction technologies are not simply neutral extensions of the body but crystallizations of prior bodily techniques, meaning the microphone itself is designed around specific, predefined expectations of how a vocal body should behave. In this sense, the microphone is not simply a passive extension of the actor's voice, but an active participant that demands specific physical adjustments. It is within this contingent negotiation that the stage actor's habits were reorganized, not through understanding the technology, but through the body's encounter with its parameters, and through a solution as simple as taking a few steps back. The stage-acting-major student did not need to know the microphone's frequency response curve. She found the solution through real-time feedback — the instructor's comments, the playback, the adjustment, and the next take. This is how the technovocalic body is formed: not through explicit knowledge of technical conditions, but through the body gradually learning to inhabit those conditions, guided by feedback rather than understanding, in ways that remain only partially visible even to the practitioner themselves. In this sense, the studio's technical parameters function as what Davis calls a materialized action, which "recognizes technological efficacy (technologies do things) but rejects the idea of technological agency" and the institutional power is not imposed on the users with a determined effect but only uncertainty.[80] In other words, the studio encodes institutional values and professional standards in the affordances and constraints of the equipment itself, and the practitioner who adjusts their body to meet those parameters is, without necessarily knowing it, being shaped by the institutional logic that designed them. The interactions reveal that the development of professionalism in the Chinese voice-acting industry is not solely attributable to individual talent. Beyond students' and instructors' accounts of becoming professionals, professionalism requires a strong ability to perform emotions well as an artistic

[80] Davis, 57-58.

achievement. Professionalism is produced under specific technological and institutional conditions, which bodies are trained to be compatible with.

How to Catch the Digital Frankenstein

At the end of each day's training, the audio engineer would organize the recording clips of each participant and post them to the group. When I heard my own recording for the first time, I was impressed by the quality of my performance. The "professional feeling" wasn't that I was a good performer or that my voice sounded clean after comping. Rather, I was surprised at how complete the three-minute clip sounded. But the fact is, these clips were not presented as complete recordings. During the sessions, I clearly remember Mr. Zhang asking me to repeat a few of these clips. So I was quite surprised to receive a full-scale track. Later, during an auditing session, I discovered the mystery of the tracking process. Under Mr. Zhang's guidance, the audio engineer selected the best performances of each part from different takes and spliced them together. In the on-screen timeline, the breathless, halting, compromising raw audio that the students kept re-recording in the studio was relentlessly sliced by the scissor tool. The audio engineer extracted the start of Take 1, the roar of Take 3, and the coda of Take 5, synthesizing them into a single, seamless recording. In doing so, the physical body that takes a few steps back in front of the microphone, searching for a place to vocalize, is completely deconstructed. In other words, when the student or audience finally hears the completed piece, it sounds unbelievably smooth, natural, and full of energy. The technology creates perfect emotional tension while erasing the voice actor's conditional labor in the studio and the friction caused by dissonance. Upon repeated listening to these finished edits, I realized that if I hadn't been in the studio observing the entire recording process in person, simply listening would have been completely unaware of the compromises between capturing the digital signal and the physical body.

This discovery revealed the last and most opaque level in the formation of the technocalic body. In the first two cases, the voice actor perceived, at least to some extent, that they were adjusting to the microphone by practicing vocalizations and varying the distance from it. But here, the final form of the professional voice was constructed in their complete absence. They don't know which takes have been chosen, which parts have been modified, or what technical processing their voices have undergone before the final listener hears them. This spliced-out voice is a Digital Frankenstein, a body assembled from multiple discrete moments of performance that never existed in that full form. Unlike the fictional monster, this digital monster's sutures are completely hidden by the production technology. This seamlessness is not accidental, since it is well documented in Helmreich's account of transduction with the case of the submarine soundscape. According to him, transduction produces "a sense of effortless presence of 'immersion'" to allow scientists inside Alvin to have "the possibility of us imagining ourselves immersed in a submarine soundscape."[81] He further suggests that transduction is the result of work, which "produces a sense of seamless presence, presence we should not take for granted but rather should inquire into as itself a technical artifact."[82] Here, the digital Frankenstein operates similarly. The compiled track does not sound like an assembly of fragments because the transductive labor of selection, editing, and mixing, aided by studio equipment, has been carried out so thoroughly that its traces are no longer audible. What the listener hears and what the practitioner hears when they first receive the finished file is not the contingent, compromised process of its production. Still, the professional standard is expected to be met in the technological environment.

How to uncover the digital monster has also become a methodological debate being explored in sound studies. When the material production process of sound is hidden by

[81] Stefan Helmreich, "Transduction," in *Keywords in Sound*, ed. David Novak and Matt Sakakeeny (Durham, NC: Duke University Press, 2015), 225.
[82] *Ibid.*, 226.

technology, simply listening is no longer enough. After a progressively deeper dialogue with STS, sound studies have moved beyond a focus on what can be heard and placed an emphasis on the interaction between the material fusion of modern sound and its multisensory action.[83] Thus, an acoustic epistemology for voice acting cannot remain at the merely auditory level, but rather needs to evolve into a studio acoustemology, "one that knows through an ongoing cumulative and interactive process of participation and reflection," especially those that take place inside the studio.[84] It is only through corporeal participation, visual observation, and the dismantling of the physical material in which the body participates that one can access the interior of the production process and see the overlooked or intentionally hidden sutures.

[83] Trevor Pinch and Karin Bijsterveld, "New Keys to the World of Sound," introduction to *The Oxford Handbook of Sound Studies*, ed. Trevor Pinch and Karin Bijsterveld (Oxford: Oxford University Press, 2012), 10, https://doi.org/10.1093/oxfordhb/9780195388947.013.0010.
[84] Steven Feld, "Acoustemology," in *Keywords in Sound*, ed. David Novak and Matt Sakakeeny (Durham: Duke University Press, 2015),13-14.

Conclusion: A Multimodal Future

By theorizing the studio as "acoustic infrastructure" and the body for voice acting as a "technovocalic body," this research contributes to STS by challenging the flat-ontology assumptions present in strands of New Materialism that assign equal agency to both human and nonhuman actants. On the one hand, this research recognizes the significance of relationality largely established by new materialism, notably, "humans, bodies, and non-human entities do not pre-exist but instead come into being 'only in and through relationships'."[85] Similarly, the idealized professional voice actor comes into being only when the studio, as the production environment, and the body trained for this context coexist and interact. By arguing against anthropocentrism, the relationship between production technologies and the human body lies in entanglement rather than linear connection.[86] But on the other hand, this does not mean acknowledging a possible equality between humans and nonhumans. While scholars such as Jane Bennett have persuasively argued for recognizing the vibrancy of matter, insisting that nonhuman entities possess a capacity to act that exceeds their instrumental use by humans, namely their agential power, this statement risks obscuring the structurally dynamic relationships between humans and the world they perceive with modes of doing.[87] The case of voice acting I present here makes this asymmetry especially visible. Voice actors do not interact with recording technologies as partners in a symmetrical assemblage. Instead, they adapt their bodies to meet technical parameters whose logic remains largely opaque to them. At the same time, the studio's institutional arrangements actively manage the boundaries of what they are expected to know. By introducing cases from my fieldwork in the program, this research aims to reveal the tension more vividly and intimately.

---

[85] Anna Puzio, "The Entangled Human Being: A New Materialist Approach to Anthropology of Technology," *AI Ethics* 5 (2025): 2342, https://doi.org/10.1007/s43681-024-00537-z.
[86] *Ibid.*, 2344.
[87] Abrahamsson et al., 6.

Focusing on a Chinese case, this research brings empirical and analytical attention to an understudied domain of contemporary Chinese popular culture. While the growth of China's "ear economy" has generated significant industry commentary and market analysis in recent years, academic scholarship on the production infrastructure and labor conditions of Chinese voice acting remains remarkably thin. Existing studies tend to approach voice acting as a supplementary element for fan culture, adaptation production, or media circulation, rather than as a site of cultural production worthy of notice. By foregrounding the institutional structures of the voice acting industry, particularly the program for young career starters, my thesis contributes to a more grounded understanding of how professionalism is constructed, transmitted, and legitimized in the emerging but poorly studied media industry. It also intervenes in broader conversations about creative labor in China's platform-driven cultural economy, where institutional legitimacy often depends on navigating opaque standards set by organizations whose authority derives partly from their control over production infrastructure.

Although the context of this research is unique to China, namely a commercially oriented training program that fills the gap left by the absence of conventional academic pathways, I do not intend to claim that my theorization of acoustic infrastructure and the technovocalic body is a China-specific framework. As previous chapters unfold, the studio as an acoustic infrastructure is also embedded within a large network of digital media production. As the technical standards and industrial logic are in transnational circulation, the theoretical intervention here is also applicable to an international context. Thus, by introducing a Chinese case into the field of media and infrastructure studies, it still has the potential to engage with a broader discussion of power dynamics in technologically mediated labor.

Lastly, this thesis makes a methodological contribution by demonstrating how sensory studies and autoethnography can be productively combined to study technologically mediated vocal practice. By redirecting analytical attention from the reception and circulation of sound to its material production, this research treats sound and listening as an epistemological angle to perceive, participate in, and intertwine with the world. By taking the Chinese case as my primary inquiry, this study asks how sound can be studied further. Studio acoustemology offers a research pathway into the recording studio production process by focusing on how practitioners perceive and adapt to their technological environment through (the production of) sound. Unlike studio studies, which are largely based on music recordings, the autoethnographic approach allows me, as the researcher, and my readers to encounter opaque body-technology interactions from the inside out. In this case, the studio is not simply where voice acting occurs; it is where the very meaning of voice, work, and audibility is continually reimagined, an insight that extends beyond this case and invites broader rethinking of how media practices are materially and socially achieved.

When direct recordings are unavailable and descriptions of sound focus on visualizing the space and scene, what are the possible approaches to studying and preserving sound-based materials? This question drives my attempt here to investigate and present my fieldwork creatively. According to Sarah Pink's multimodal framework, differentiated senses are not separate from one another when perceiving the world and environment, nor do they assign specific modes to categories.[88] In other words, in my research, sound is not an exclusive category limited to auditory experiences. Instead, considering multimodality, it can be presented through other sensory modes, acknowledging the importance and ability of drawings. This conclusion, then, is not only a reflection on what drawing can do for ethnography in general, but also an account of how a specific drawing strategy allows me to

[88] Sarah Pink, "Multimodality, Multisensoriality and Ethnographic Knowing: Social Semiotics and the Phenomenology of Perception," *Qualitative Research* 11, no. 3 (2011): 266, https://doi.org/10.1177/1468794111399835.

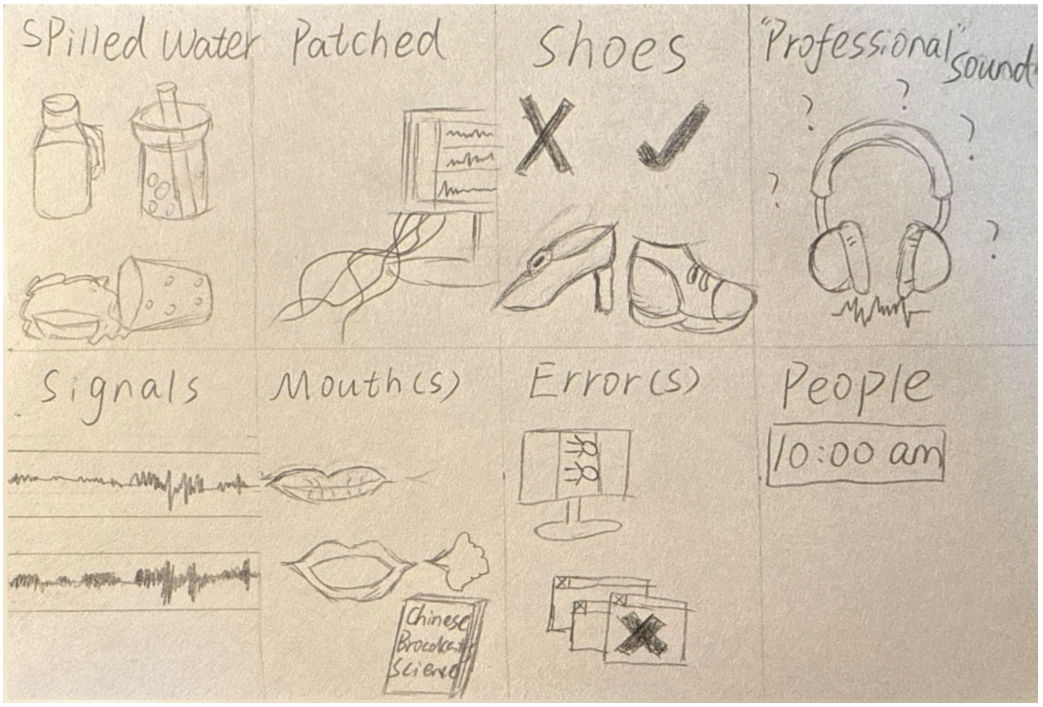
render the sonic textures and embodied protocols of the Chinese audio industry in ways that writing alone could not.

In anthropology, drawings, such as diagrams, are widely used to convey complex ideas through abstraction, simplification, and detail. These drawings usually emerge from analysis rather than those that constitute them and negotiate between what can and cannot be seen in visuals.[89] However, as Douglas-Jones suggests, drawings are also analytical tools that "are more felt than thought, that emerge from conceptual proximities, observed elisions, conflicts, gaps, and erasures" rather than merely explaining ethnographic material or generalizing.[90] As she explains, the analytical drawings help organize materials in a non-linear structure, which makes connections between practices without imposing a predetermined configuration of thought, and then supports a comprehensive understanding of the experience for further writing. Her framework for drawing, though encouraging a discursive and decentralized approach, still serves as a tool for future rational knowledge production that is not burdened by the evaluation of artistry.[91] In my own project, this mode prompts me to hold together scattered observations about the studio, vocal warm-ups, and the rules governing practitioners' bodies, resisting the linear flow of written ethnography.

Compared to illustrative diagrams, the images in *Things that Art* do not provide a clear explanation and are presented in a more artistic way. Lochlann Jain's hand-drawings with accompanying word explanation shed light on our interests in "the energy of the varied and discordant claims alongside one another rather than any settled determination," which then reveal gaps that allow viewers to actively engage with, find, and create meaning in the artwork.[92] By drawing something under the category pairing with textual labels, Jain's work of art does not hold one type of truth, but invites readers to interpret the dissonance or

[89] Rachel Douglas-Jones, "Drawing as Analysis: Thinking in Images, Writing in Words," in *Experimenting with Ethnography: A Companion to Analysis* (Durham, NC: Duke University Press, 2021), 97.
[90] *Ibid*.
[91] *Ibid.* 105.
[92] Lochlann Jain, *Things That Art: A Graphic Memoir* (Toronto: University of Toronto Press, 2019), 43.

collision. Her work inspires me to take the chance and draw out something that may not be included in my body text, but would be interesting to remind my readers that there are other details to explore beyond my full writing.



(Figure 9: the drawing I made on a piece of card, inspired by Jain's works)

Figure 9 shows my first attempt to incorporate drawings in a multimodal project. For the sake of more individualized interpretations, I will introduce only one section of the drawing and my approach to organizing it. For example, in the third section of the first row in Figure 9, I drew two pairs of shoes. The drawing carries layered meanings that resist a single reading. At the most literal level, it documents a studio rule: during recording sessions, practitioners are required not to wear clothing or jewelry that rattles or makes noise, including high heels. The shoes, in this sense, index a sonic protocol: what the microphone must not hear. But at the same time, the two pairs also echo the stereotypical male/female icons hung on restroom doors. Once read through that frame, the image opens onto a different claim that within the Chinese audio industry, women practitioners across different roles receive noticeably less attention than their male counterparts, especially male voice actors.

Following Jain, I leave the dissonance visible rather than resolving it in a caption. By doing so, I trust readers to sit with the collision between a technical rule about silencing the body and a gendered hierarchy about whose voice gets heard.

The drawing strategy gestures toward what a multimodal future might look like for ethnographies of sound. When the sounds themselves cannot be played back directly, and when prose risks flattening the embodied, spatial, and gendered dimensions of sonic practice, drawings offer a way to keep those dimensions in tension on the page. I do not claim that visual materials can substitute for sound, nor that my own hand can render what a microphone captures. What I do claim is that the page, being treated as a multimodal surface, can hold more of the fieldwork than the paragraph alone. If Pink's framework asks us to stop sorting experience into isolated sensory modes, then the closing proposition of this project is an invitation: to keep drawing, collaging, and otherwise visualizing sound not as illustration, but as a mode of thinking with the ear through the eye.

Appendix

The following audio examples are provided as supplementary materials to illustrate the sonic textures discussed in the analysis. The audio files presented here are from my fieldwork on the training program. They are uploaded to SoundCloud as public.

1. Audio Example 1: the practice piece recorded in the recording booth. https://on.soundcloud.com/war3j2IACaVHXYw6f2
2. Audio Example 2: the practice piece recorded in the classroom during the final showcase of the program. https://on.soundcloud.com/BSYmeiu7Q6ddcxW7DK

Bibliography

Abrahamsson, Sebastian, Filippo Bertoni, Annemarie Mol, and Rebeca Ibáñez Martín. "Living with Omega-3: New Materialism and Enduring Concerns." *Environment and Planning D: Society and Space* 33, no. 1 (2015): 4–19. https://doi.org/10.1068/d14086p.

ACGx. "Cong Muhou dao Taiqian: Zhongguo Wangpei de Shangye Hua Zhi Lu 从幕后到台前：中国网配的商业化之路 From Behind the Scenes to the Forefront: The Commercialization of Chinese Online Voice Acting." *Huxiu* 虎嗅, November 14, 2018. https://m.huxiu.com/article/271753.html.

Alburger, James R. *The Art of Voice Acting: The Craft and Business of Performing for Voiceover*. 7th ed. New York: Focal Press, 2023.

Baron, Jaimie. "Ventriloquizing Obama, or, the Ethics of Archival Ventriloquism." In *Media Ventriloquism: How Audiovisual Technologies Transform the Voice-Body Relationship*, edited by Jaimie Baron, Jennifer Fleeger, and Shannon Wong Lerner, 199–214. Oxford: Oxford University Press, 2021.

Behrendt, Frauke. "Soundwalking." In *The Routledge Companion to Sound Studies*, edited by Michael Bull, 249-57. London: Routledge, 2018.

Chang, Heewon. *Autoethnography as Method*. London and New York: Routledge, Taylor & Francis Group, 2016.

Chen, Jasmine Yu-Hsing. "Beyond Words: Navigating Aural Fantasies in Chinese Boys' Love Audio Drama." M/C Journal 27, no. 2 (2024). https://doi.org/10.5204/mcj.3033.

Chion, Michel. *Audio-Vision: Sound on Screen*. Translated by Claudia Gorbman. New York: Columbia University Press, 2019.

Chu, Warren. “Bridging Linklater Voice and Meisner Acting within the Chinese Classroom.” *Voice and Speech Review* 19, no. 1 (2022): 4–16. https://doi.org/10.1080/23268263.2022.2143604.

Davies, Rod. “An Autoethnography of Performance in the Vocal Booth.” In *The Routledge Companion to Music, Autoethnography, and Reflexivity*, edited by Peter Gouzouasis and Christopher Wiley, 88–105. London: Routledge, 2024.

Davis, Jenny L. “Politics and Power.” In *How Artifacts Afford: The Power and Politics of Everyday Things*, 45–61. Cambridge, MA: MIT Press, 2020. https://doi.org/10.7551/mitpress/11967.003.0005.

Douglas-Jones, Rachel. “Drawing as Analysis: Thinking in Images, Writing in Words.” In *Experimenting with Ethnography: A Companion to Analysis*, 94–105. Durham, NC: Duke University Press, 2021.

Edwards, Paul N., Geoffrey C. Bowker, Steven J. Jackson, and Robin Williams. “Introduction: An Agenda for Infrastructure Studies.” *Journal of the Association for Information Systems* 10, no. 5 (2009): 365–74. https://doi.org/10.17705/1jais.00200.

Eidsheim, Nina Sun. *Sensing Sound: Singing and Listening as Vibrational Practice*. Durham: Duke University Press, 2015. https://doi.org/10.1215/9780822374695.

Feld, Steven. “Acoustemology.” In *Keywords in Sound*, edited by David Novak and Matt Sakakeeny, 12–28. Durham: Duke University Press, 2015.

Harris, Ben. “Introduction to Acoustics.” In *Home Studio Setup: Everything You Need to Know from Equipment to Acoustics*, 1–14. New York: Routledge, 2009. https://doi.org/10.1016/B978-0-240-81134-5.00001-2.

Helmreich, Stefan. “Transduction.” In *Keywords in Sound*, edited by David Novak and Matt Sakakeeny, 222–231. Durham, NC: Duke University Press, 2015.

Hsieh, Jennifer C. “Sound Studies and East Asian STS: Revisiting the Sonic Legacy of Trevor Pinch.” *East Asian Science, Technology and Society: An International Journal* 19, no. 1 (2025): 89–104. https://doi.org/10.1080/18752160.2024.2359252.

Huang, Ying. *Ear Economy: China Audio Streaming Programs Study*. Singapore: Palgrave Macmillan, 2024.

Hu, Tingting, Jing Jin, and Lin Liao. “A Transmedia ‘Third’ Space: The Counterculture of Chinese Boys’ Love Audio Dramas.”*Asian Studies Review* 47, no. 4 (2023): 836–854. https://doi.org/10.1080/10357823.2023.2211223.

“Huanhuan Zhuanye Jingshen: Yanyuan Suzao Juese Buneng Pinpin Yong Peiyin 呼唤专业精神:演员塑造角色不能频频用配音 Calling for Professionalism: Actors Should Not Frequently Use Dubbing to Portray Characters.” *Guangming Daily* 光明日报, April 23, 2023. https://www.chinanews.com/cul/2023/04-23/9995223.shtml.

Ingold, Tim. “On Weaving a Basket.” In *The Perception of the Environment: Essays on Livelihood, Dwelling and Skill*, 426–438. London: Routledge, 2021. https://doi.org/10.4324/9781003196662.

Jain, Lochlann. *Things That Art: A Graphic Memoir*. Toronto: University of Toronto Press, 2019.

Kielman, Adam. “Sonic Infrastructures, Musical Circulation and Listening Practices in a Changing People’s Republic of China.” *Sound Studies* 4, no. 1 (2018): 19–34. https://doi.org/10.1080/20551940.2018.1499067.

Larkin, Brian. *Signal and Noise: Media, Infrastructure, and Urban Culture in Nigeria*. Durham, NC: Duke University Press, 2008.

Larkin, Brian. “The Politics and Poetics of Infrastructure.” *Annual Review of Anthropology* 42 (2013): 327–43. https://doi-org.proxy.lib.umich.edu/10.1146/annurev-anthro-092412-155522.

Levy-Landesberg, Hadar, and Xuenan Cao. “Anchoring Voices: The News Anchor’s Voice in China from Television to AI.” *Media, Culture & Society* 47, no. 2 (2025): 229–251. https://doi.org/10.1177/01634437241270937.

Li, Jie. “Revolutionary Echoes: Radios and Loudspeakers in the Mao Era.” *Twentieth-Century China* 45, no. 1 (2020): 25–45. https://dx.doi.org/10.1353/tcc.2020.0006.

Lin Feng, “Voicing the West: Dubbing Artists and Invisible Labor Behind the Socialist Screen,” in *Film History and Screen Culture in and beyond Greater China*, ed. Lin Feng (London: Routledge, 2025), 74, https://doi-org.proxy.lib.umich.edu/10.4324/9781003378839.

Liu, Tingting, and Zishan Lai. “Eroticism and Design Workshops in Game Interfaces: Exploring a Mobile Otome Game for Chinese Women.” *Games and Culture*, December 22, 2024, 1–21. https://doi.org/10.1177/15554120241306788.

Lupton, Deborah. “How Do Data Come to Matter? Living and Becoming with Personal Data.” *Big Data & Society* 5, no. 2 (2018): 1–11. https://doi.org/10.1177/2053951718786314.

Mattern, Shannon. “Maintenance and Care.” *Places Journal*, November 2018, 1–31. https://doi.org/10.22269/181120.

Meintjes, Louise. “The Recording Studio as Fetish.” *Sound of Africa!: Making Music Zulu in a South African Studio*, 265–82. Durham: Duke University Press, 2003.

Nozawa, Shunsuke. “Ensoulment and Effacement in Japanese Voice Acting.” In *Media Convergence in Japan*, edited by Patrick W. Galbraith and Jason G. Karlin, 169–199. Ann Arbor: Kinema Club, 2016.

Parks, Lisa. “Stuff You Can Kick: Toward a Theory of Media Infrastructures.” In *Between Humanities and the Digital*, edited by Patrik Svensson and David Theo Goldberg,

355-373. Cambridge, MA: MIT Press, 2015. https://doi-org.proxy.lib.umich.edu/10.7551/mitpress/9465.003.0031.

Phan, Phoebe. "The Materiality of the Digital and the Gendered Voice of Siri." *Transformations* 29 (2017): 23–33. https://transformationsjournal.org/wp-content/uploads/2021/02/Phan_Transformations29.pdf.

Pinch, Trevor, and Karin Bijsterveld. "New Keys to the World of Sound." Introduction to *The Oxford Handbook of Sound Studies*, edited by Trevor Pinch and Karin Bijsterveld, 3–36. Oxford: Oxford University Press, 2012. https://doi.org/10.1093/oxfordhb/9780195388947.013.0010.

Pinch, Trevor, and Karin Bijsterveld. "Sound Studies: New Technologies and Music." *Social Studies of Science* 34, no. 5 (2004): 635–648. https://doi.org/10.1177/0306312704047615.

Pink, Sarah. *Doing Sensory Ethnography*. 2nd ed. London: SAGE Publications, 2015. https://doi.org/10.4135/9781473917057.

Pink, Sarah. "Multimodality, Multisensoriality and Ethnographic Knowing: Social Semiotics and the Phenomenology of Perception." *Qualitative Research* 11, no. 3 (2011): 261–276. https://doi.org/10.1177/1468794111399835.

Porcello, Thomas. "Speaking of Sound: Language and the Professionalization of Sound-Recording Engineers." *Social Studies of Science* 34, no. 5 (2004): 733–758. https://doi.org/10.1177/0306312704047328.

Pras, Amandine, Catherine Guastavino, and Maryse Lavoie. "The Impact of Technological Advances on Recording Studio Practices." *Journal of the American Society for Information Science and Technology* 64, no. 3 (January 16, 2013): 612–26. https://doi.org/10.1002/asi.22840.

Puzio, Anna. “The Entangled Human Being: A New Materialist Approach to Anthropology of Technology.” *AI Ethics* 5 (2025): 2339–2356. https://doi.org/10.1007/s43681-024-00537-z.

Ramati, Ido. “Algorithmic Ventriloquism: The Contested State of Voice in AI Speech Generators.” *Social Media + Society* 10, no. 1 (2024): 1–10. https://doi.org/10.1177/20563051231224401.

Rice, Tom. “Ethnographies of Sound.” In *The Routledge Companion to Sound Studies*, edited by Michael Bull, 239-48. London: Routledge, 2018.

Star, Susan Leigh, and Geoffrey C. Bowker. “How to Infrastructure.” In *Handbook of New Media: Social Shaping and Social Consequences of ICTs*, updated student ed., edited by Leah A. Lievrouw and Sonia Livingstone, 230–245. London: SAGE Publications, 2010. https://doi.org/10.4135/9781446211304.n13.

Schmidt Horning, Susan. *Chasing Sound: Technology, Culture, and the Art of Studio Recording from Edison to the LP*. Baltimore: Johns Hopkins University Press, 2013.

Siegert, Bernhard. “Doors: On the Materiality of the Symbolic.” Translated by John Durham Peters. *Grey Room*, no. 47 (2012): 6–23. https://doi.org/10.1162/GREY_a_00067.

Sterne, Jonathan. “The MP3 as Cultural Artifact.” *New Media & Society* 8, no. 5 (2006): 825–842. https://doi.org/10.1177/1461444806067737.

Sterne, Jonathan. “Space within Space: Artificial Reverb and the Detachable Echo.” *Grey Room*, no. 60 (Summer 2015): 110–131. https://www.jstor.org/stable/43832233.

Thompson, Emily. “Sound, Modernity and History.” In *The Sound Studies Reader*, edited by Jonathan Sterne, 117–129. New York: Routledge, 2012.

Verma, Neil. *Theater of the Mind: Imagination, Aesthetics, and American Radio Drama*. Chicago: University of Chicago Press, 2012.

Wang, Ting 王婷. "'Erduo Jingji' Pengbo Fazhan '耳朵经济'蓬勃发展 The Booming Development of the 'Ear Economy'." *Minsheng Zhoukan* 民生周刊, June 9, 2025. https://paper.people.com.cn/mszk/pc/content/202506/09/content_30082226.html.

Watson, Allan. *Cultural Production in and Beyond the Recording Studio*. New York: Routledge, 2014. https://doi.org/10.4324/9780203728260.

Weidman, Amanda. "Voice." In *Keywords in Sound*, edited by David Novak and Matt Sakakeeny, 232–245. Durham: Duke University Press, 2015.

Wu, Hongyi 吴弘毅. *Shiyong Boyin Jiaocheng 1* 实用播音教程1 Practical Broadcasting Tutorial, vol. 1. Beijing: Communication University of China Press 中国传媒大学出版社, 2002.

Xia Lei 夏磊. "走进'替'声的世界，讲述真实的配音行业." *TED*中文. 2016. https://www.ted.com/podcasts/tedinchinese-transcripts-xialei2016.

Yandong 彦东. "Zhongguo Shengyou de Nixi Zhi Lu 中国声优的逆袭之路 The Comeback Path of Chinese Voice Actors." *Huxiu* 虎嗅, June 20, 2017. https://m.huxiu.com/article/200600.html.

Zheng, Xiaofeng. "One Year of Soviet Film Translation Work: Several Issues in the Artistic Creation of Dubbed Films 一年来苏联影片的译制工作:翻译片在艺术创造上的几个问题 yinian lai sulian yingpian de yizhi gongzuo: fanyi pian zai yishu chuangzao shang de ji ge wenti." *Guangming Daily*, March 17, 1951.