

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Listening Through the Firewall: A Sonic Narrative of Communication Between Taiwan and Mainland China

Permalink

<https://escholarship.org/uc/item/5qq7p4hx>

Author

Plovnick, Sarah E

Publication Date

2024

Peer reviewed|Thesis/dissertation

Listening Through the Firewall:
A Sonic Narrative of Communication Between Taiwan and Mainland China

By

Sarah E Plovnick

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Music

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jocelyne Guilbault, Chair

Professor Andrew F. Jones

Professor Daniel Fisher

Summer 2024

Abstract

Listening Through the Firewall:
A Sonic Narrative of Communication Between Taiwan and Mainland China

by

Sarah E Plovnick

Doctor of Philosophy in Music

University of California, Berkeley

Professor Jocelyne Guilbault, Chair

This dissertation adopts an ethnographic approach to understand the complex political relations between Taiwan and mainland China, focusing on how sound and audio communication have been central in maintaining continuous dialogue across the Taiwan Strait even amidst political tensions and communication restrictions. It demonstrates that throughout 75 years of cross-strait history, people in Taiwan and mainland China have continuously found ways to speak and listen to one another. Since the Kuomintang government's retreat to Taiwan in 1949 at the end of the Chinese Civil War, communication across the Taiwan Strait has been heavily restricted. Both Taiwan and mainland China's governments maintained highly regulated media environments through the 1980s, and mainland China's Great Firewall continues to impose strict controls on online communication today. Amidst this restricted environment, people have mobilized various audio communication technologies—from loudspeakers and radio broadcasts to social media and videogames—to project a steady stream of sound across the Taiwan Strait. While discussions of cross-strait politics often emphasize formal diplomatic negotiations and political tension, this research instead centers sound and voice communication as distinct perspectives through which to explore how individuals communicate beliefs and navigate differences as they go about their daily lives.

A recurring theme in this dissertation is the concept of *regulatory latency*, which describes how the development and implementation of regulatory mechanisms often lag behind those of communication technologies. This latency leads to unregulated gaps that can be temporarily leveraged by individuals for open exchange in highly restricted media environments. I build on the concept of regulatory latency to demonstrate that since the 1950s, individuals in Taiwan and mainland China have continuously employed audio technologies to communicate outside of the bounds of the latest government regulations. I start with loudspeakers, which were used from the 1950s on both sides of the Taiwan Strait not only to transmit cross-strait propaganda, but also for informal exchange. Radio broadcasts similarly provided a rare space for cross-strait exchange; I specifically focus on the transitional period from the late 1980s through the early 2000s during which Taiwanese radio hosts and mainland Chinese listeners formed sustained relationships through multimedia exchanges surrounding these broadcasts. Analyzing these exchanges

encourages a reconceptualization of broadcast radio as multimedia and multidirectional, while also revealing shifts in what I call the cross-strait socio-political imaginary.

More recently, I consider online exchanges through a case study of the briefly popular audio social media app Clubhouse. I highlight how Clubhouse users engaged in live voice communication to escape the “echo chambers” that are common on contemporary social media, forming affective connections and empathizing with people whose beliefs differ from their own. I also explore live-voice communication in online videogames. In contrast to associations of videogames with the promotion of violence, I demonstrate how informal, relaxed interactions in the context of online play provide a site through which Taiwanese and mainland Chinese gamers can collectively confront cultural, linguistic, and political differences. Taken together, these case studies reveal situations in which individuals find ways to communicate despite restrictions, navigating their differences by speaking and listening to one another.

In memory of my grandmother, Sylvia G. Krakow, who earned her doctorate in 1964 and paved the way for many future generations of women.

Contents

Contents	ii
List of Acronyms	iii
Note on Language Usage	iv
Acknowledgments	v
1. Introduction: From Walls of Sound to Firewalls	1
2. The Golden Era of Cross-Strait Exchange: Multidirectional Broadcast Radio from the Late 1980s through Early 2000s	13
3. Out of the Echo Chamber: Cross-Strait Exchange During the Clubhouse Moment	41
4. Negotiating Political Differences through Online Play	61
5. Continuing the Cycle: Audio Communication and Regulation in the Age of AI	81
Works Cited	108

List of Acronyms

AI - Artificial Intelligence

ASR - Automatic Speech Recognition

CBS - Central Broadcasting Station

CCP - Chinese Communist Party

CNN - Convolutional Neural Network

DPP - Democratic Progressive Party

DSP - Digital Signal Processing

GPU - Graphics Processing Unit

KMT - Kuomintang

KOL - Key Opinion Leader

PLA - People's Liberation Army

PRC - People's Republic of China

ROC - Republic of China

RTI - Radio Taiwan International

TSMC - Taiwan Semiconductor Manufacturing Company

VoIP - Voice over Internet Protocol

VPN - Virtual Private Network

Vtuber - Virtual YouTuber

WoW - World of Warcraft

Note on Language Usage

This dissertation is written in English, with Mandarin Chinese translations provided for selected proper nouns. For content originally in Chinese (including excerpts from publications, passages from interviews, and song lyrics), I provide the English translation and the original Chinese text or transcript. All translations are mine unless otherwise noted. Translations use traditional Chinese characters, since this is what I was trained on and what is commonly used in Taiwan. There are a few exceptions where simplified characters are used, particularly for online comments originally written in simplified characters. For Mandarin place names and personal names, I use the Hanyu Pinyin transliteration system unless there is another commonly accepted English spelling or an expressed desire for an alternate spelling (e.g. Tsai Ing-wen, not Tsai Ying-wen). For the names of Taiwanese people, I follow the convention of placing a hyphen between given names (e.g. Ying-wen).

Acknowledgments

This dissertation is as much a collective effort as it is my own work, borne from many fascinating and memorable conversations with people in Taiwan, Berkeley, and online. Some of these people are named throughout this dissertation, while others remain anonymous. I am deeply grateful to each and every person who took the time to meet and share a small portion of their work and lives with me. I appreciate their insightful answers and their patience even when I asked too many questions or took too long to understand.

I would like to express my deepest gratitude to my advisor, Professor Jocelyne Guilbault. The wide scope of this dissertation is the result of Jocelyne's encouragement to think beyond disciplines and the example she provided to strive continuously to understand new ideas. Jocelyne taught me that even the smallest details merit attention, with a single word powerful enough to change the meaning of an entire argument. Her ongoing attention, support, and accountability have been invaluable throughout my doctoral journey.

I wish to extend my heartfelt thanks to the other members of my committee, Professors Andrew Jones and Daniel Fisher. I am indebted to Andrew for his insightful and candid feedback, as well as for introducing me to many helpful contacts within the academic community in Taiwan. I learned a great deal from his seemingly limitless expertise of Taiwan's language, culture, and history, and he was always ready with the perfect literary reference to complement any argument. Coursework with Daniel in media anthropology and sound studies was instrumental in informing many of the ideas developed in this dissertation. His feedback on my work has pushed me to develop my ideas in dialogue with cutting edge literature. I also want to thank Professor Tom Gold, an unofficial committee member, who provided a thorough introduction to scholarship in Taiwan studies and connected me to his vast network in the Taiwan studies and policy communities.

I feel incredibly grateful to have been a part of the UC Berkeley music department. I was privileged to have the opportunity to take courses with many brilliant professors during my time at UC Berkeley, which supplied the essential foundation and training necessary to complete my PhD. Thank you to Ben Brinner, Nick Matthews, Maria Sonevytsky, Tom Porcello, Myra Melford, and Carla Brunet for providing a welcoming space to grow. This community would not be what it is without the ethnomusicology, musicology, and composition graduate students, who actively and intentionally create a positive and welcoming environment. I am grateful for the generational knowledge passed down in Morrison 107, the solidarity from classmates during challenging seminars, and the many spontaneous discussions that led to new ideas. In particular, I want to thank the members of my cohort, Ryan Gourley and Jon Turner, who have been there since the beginning with unwavering friendship and camaraderie through the ups and downs of graduate school. Our countless slices of pizza at Artichoke, coffees at Strada, happy hours, wine Wednesdays, and more transformed the past seven years from tolerable to truly enjoyable. In my final year, the Berkeley Arts and Ethnography (BAE) writing group came to my rescue with our weekly coffee meetings. Ryan as well as Chris Chan, Jon Wu, and Junius Brown were willing to read all of my worst drafts, providing crucial feedback that propelled my dissertation to its current form. Thank you for joining me on my search for the best cafe in Berkeley and for always having the appropriate meme ready for any occasion. Thanks also to BAE's newest members Max Jefferson and Caitlin Romtvedt, who have recently invigorated our group with

fresh energy. And a special shout-out to our youngest member, Ornette—I am looking forward to reading his work in 2042!

This dissertation would not have been possible without the generous and welcoming support I received from colleagues and collaborators in Taiwan. In particular, I am grateful to the faculty and students at the Graduate Institute for Musicology at National Taiwan University for welcoming me into their community. Wilson Yick-sau Lau was the first person I met from this community back in 2019, and he facilitated many essential connections for my research and life in Taiwan. Professor DJ Hatfield kindly served as a local advisor for my project. DJ alongside Shura Taylor and Seung Im Seo contributed transformative feedback on an early presentation of Chapter 4 during a very fun panel at AAS-in-Asia 2023. Professor Fumitaka Yamauchi was also instrumental in providing commentary on early versions of this work.

So much of this project was dependent on knowing and finding people with very specific, relevant expertise across disciplines. I am indebted to the many people who facilitated this process. I want to especially recognize Professor Yi-hsuan Eric Yang, who showed interest and support for my ideas even before I arrived in Taiwan and was generous in introducing me to his network of established researchers from Taiwan's AI community. Chapter 5 would not have been possible without his connections and insights. Radio Taiwan International and Rachel Luo in particular kindly contributed access to their fascinating archive and provided important introductions to members of the RTI community. These resources served as a rich foundation for Chapter 2. Professor Lawrence Yang offered many useful introductions while also expressing consistent support for my work. Mark Hsiang-Yu Feng contributed insightful feedback and key connections for Chapter 4. I also want to acknowledge Dayton Lekner, who created an invaluable global network of scholars when he established the East Asia Sound Studies working group. Its members (the "resonators") provided essential feedback on an early version of Chapter 3, and their research supplied much inspiration and background information throughout the dissertation. Among this group, Yusheng Lei offered helpful introductions to many insightful mainland Chinese internet users. Thanks also to Yuanqi Lyu for showing continuous enthusiasm for this project.

When I started studying Mandarin Chinese in 2018, I never imagined I would be able to use the language to carry out even a simple conversation, let alone a doctoral research project. I want to express my appreciation to language instructors Mingzhe Zheng, Damien Donnelly, and Linda Zhang at Berkeley as well as the many talented teachers and staff at the International Chinese Language Program in Taipei. Their continuous patience and consistent corrections made this work possible. Learning a language requires a lot of practice, and for the last four years Mingying Tsai has listened to me stumble through both new and forgotten vocabulary on a weekly basis. I am so grateful to her for being such a consistent language exchange partner, whether in Taiwan or online. Thank you for being an amazing friend and for making Taiwan feel a bit more like home. Our regular discussions continue to be a highlight of my week.

I want to acknowledge the generous funding organizations that contributed support for travel and language study, as well as the time and space to think through ideas. These include Fulbright-Hays DDRA, the Center for Chinese Studies at Taiwan's National Central Library, Berkeley's Center for Chinese Studies, Berkeley GIAS, Taiwan's Ministry of Education, Taiwan's Ministry of Foreign Affairs, and FLAS. I especially want to recognize Yi-chun Yeh at Taiwan's National Central Library for his expertise in coordinating the visiting scholar program, as well as the staff at Fulbright Taiwan for facilitating the visa process and organizing valuable

networking events. My Fulbright cohort in Taiwan also served as a wonderful community to fall back on while living far away from friends and family.

My interest in sound and listening extends back long before my graduate studies, and I have been fortunate to have many excellent mentors in this realm. At Washington University in St. Louis, Professors Denise Gill and Pat Burke provided guidance and expertly modeled what it meant to pursue a career in ethnomusicology. Professor Tili Boon Cuillé patiently guided me through my earliest experiences with independent research. I want to thank Bill Lenihan, Carole Lemire, and Kara Baldus-Mehrmann at WashU for supporting my music performance interest and helping me to development essential listening skills. This built upon the love of music that my childhood teachers, Jamie Saltman and Ellen Donahue-Saltman, instilled in me from a young age.

In 2019 after only one year of studying Mandarin, I arrived in Taiwan for intensive language study, not even knowing how to ask someone on the street for directions. Hsu Chin-hui (許錦惠) and Tsai Chung-jen (蔡崇仁) kindly welcomed me into their home and helped me to survive, thrive, and grow to love Taiwan. I want to extend a special thanks to my future in-laws Jean-Shou and Hudson Hsieh for their generous support throughout my studies, whether in Taiwan, Berkeley or Omaha. They are always ready to help me practice my Mandarin and Taiwanese Hokkien language skills and have encouraged my research and interest in Taiwan for many years. Thank you also to Shauyene, Dave, and Callan for their encouragement; I am looking forward to the day when Cal's Mandarin skills surpass my own.

I gratefully acknowledge my parents, Milly Krakow and Martin Plovnick, for their abundant love and support throughout too many years of education. Thank you for always encouraging me to explore the world, whether through books, music, or travel. A special thanks to my sister, Amy, my brother-in-law, Josh, and my niece, Sylvie, for always putting a smile on my face with their videocalls and visits. Finally, I want to thank my fiancé Satcher Hsieh for his unconditional love and for supporting me in more ways than are possible to count. He introduced me to Taiwan and consistently encouraged me to choose the most interesting path, even if it wasn't the easiest one. He is my first and last reader, my most honest critic, and he has fueled my writing with the best espresso in the world. This dissertation truly would not have been possible without him.

1. Introduction: From Walls of Sound to Firewalls

*I'm waiting outside the wall for you
I'm waiting for your reply
I hear your breath and your heartbeat
Namewee, "The Wall"*

我站在牆外等妳
期盼著你的回應
我聽見你的心跳還有你的呼吸
黃明志【牆外】

This dissertation begins from a seemingly simple question: how do people in Taiwan and mainland China speak and listen to one another? However, this question quickly becomes complex in the context of 75 years of political relations across the Taiwan Strait. Contemporary cross-strait relations can be traced to the end of the Chinese Civil War in 1949, when the Chinese Communist Party (CCP) established the People's Republic of China (PRC) in mainland China and the Republic of China's (ROC) Kuomintang (KMT) nationalist government fled to Taiwan, bringing with them 1.5 million political refugees.¹ Since then, cross-strait relations have been characterized by varying degrees of military tensions, diplomatic disputes, and communication restrictions. At the heart of this complex relationship lies the fundamental issue that Taiwan operates as a self-governing, independent, and democratic island nation even while the PRC continues to insist that Taiwan is part of its territory.² This conflicting stance has led to ongoing disputes and diplomatic tensions, with mainland China exerting pressure on the international community not to recognize Taiwan as a separate nation while Taiwan continues to seek greater international recognition and participation in global affairs.

Mainstream media narratives covering interactions between Taiwan and mainland China most often emphasize military drills, election interference efforts, and impending conflict; *The Economist* even went so far as to declare Taiwan "the most dangerous place on earth" (2021). And yet, these broad media narratives do not always align with the diverse experiences of individuals on the ground. This dissertation focuses on the daily lives of people on both sides of the Taiwan Strait to provide an alternative perspective of how individuals, rather than political actors, navigate social, cultural, and political differences in their everyday interactions. By taking an ethnographic approach, my research reveals that throughout 75 years of cross-strait history—even amidst political tensions and communication restrictions—people in Taiwan and mainland China have continuously found ways to speak and listen to one another.

¹ For the purposes of this dissertation, I use "mainland China" synonymously with "People's Republic of China (PRC)," and "Taiwan" synonymously with "Republic of China (ROC)." For those who arrived in Taiwan in 1949 with the KMT, continuous political tensions meant that they could not speak to relatives left behind in mainland China for a period of 40 years (Rigger 2021, 33).

² As of 2023, 62.8 percent of Taiwan's citizens identify as Taiwanese, 30.5 percent identify as both Taiwanese and Chinese, and 2.5 percent identify as Chinese (Election Study Center at National Chengchi University, 2023).

Walls of Sound

There is perhaps no better symbol of the continuous endurance of cross-strait communication than the giant loudspeaker “walls of sound” (播音牆) that were used for psychological warfare across the Taiwan Strait from the 1950s through the 1980s. During this period, loudspeaker systems in the southernmost Fujian region of mainland China blasted slogans and threats toward the ROC islands in the Taiwan Strait, which countered with their own audible psychological warfare. One of the most famous of these systems is the Beishan Broadcasting Wall, which still stands on the northernmost coast of the ROC island-group Kinmen. This structure sits near the site of the Battle of Kuningtou – one of the final battles of the Chinese Civil War in October 1949, the results of which secured the ROC’s presence in Kinmen despite Kinmen’s geographic proximity to mainland China.³ Chinese literature scholar Andrew Jones describes this formidable wall of sound:

Rising like a modernist obelisk on a cliff above the coast of the island of Quemoy (Kinmen) is a curious concrete structure. More than four stories high and symmetrically punctuated by circular portholes, the building looks like nothing so much as a gigantic loudspeaker. This is precisely what it is—perhaps the world’s largest and most imposing sound system, housing an array of forty-eight horn reflex speakers and capable of projecting sound up to twenty-five kilometers away (2020, 169).

Loudspeakers such as the Beishan Broadcasting Wall performed a strategic purpose in psychological warfare efforts while also serving as a rare point of communication for broadcasters on each side of the Taiwan Strait. For example, Jones writes that in addition to highly regulated propaganda messages, there were also occasional exchanges of friendly banter about topics such as the weather (“It’s about to rain; please take in your wash from the clothesline”) (Jones 2020, 175–76). Historian Dayton Lekner emphasizes that despite the opposing ideological content of the broadcasts, announcers would listen to the messages coming across from the other side and then adjust their own vocal broadcasting style in response. While this interactive listening was in part a strategy for broadcasters to develop more effective delivery of political propaganda, these exchanges also demonstrate a sustained, engaged, and interactive listening process, even hedging on a sense of mutual understanding between the two sides (Lekner 2022, 27).

³ Kinmen (also frequently written as Quemoy or Jinmen, and in Chinese 金門) is an ROC-governed group of islands just 6.2 miles (10 kilometers) east of the city of Xiamen in the Southern Fujian region of mainland China. It is easy to catch glimpses of Xiamen’s skyline from various points around the islands, and the ferry ride from Kinmen to Xiamen takes only 30 minutes. For more on Kinmen’s central role in cross-strait conflict since the 1950s, see Szonyi 2008. For accounts specific to Kinmen’s loudspeaker stations, see Lin 2009; Jones 2020, 169–77; Lekner 2022.



Figure 1. The Beishan Broadcasting Wall, November 2022. Photo by author.

As tensions eased in the late 1970s, loudspeaker broadcasts began to incorporate popular music to complement political propaganda. Perhaps the most famous music to be broadcast through these loudspeakers was that of Taiwanese singer Teresa Teng (鄧麗君). Teng’s voice reached mainland China through radio broadcasts as well as the loudspeakers, captivating listeners with a novel sense of intimacy to contrast the revolutionary songs that were most common in mainland China at the time.⁴ Since 1980s political and economic reforms in mainland China and the end of martial law in Taiwan in 1987, Kinmen’s place on the front lines of the Cold War has increasingly faded into the background, replaced by a developing tourism industry.⁵ Even with these changes, Teresa Teng’s voice continues to ring out from the Beishan Broadcasting Wall today, albeit at a lower volume to reach only the ears of Kinmen’s tourists rather than listeners across the Strait.⁶

⁴ For more on Teresa Teng via loudspeakers and radio broadcasts, see Chapter 2; also see Schweig 2022; Cheng 2019; Wu 2021; Jones 2020, 169–96.

⁵ Kinmen opened to tourists from Taiwan in 1993 and to tourists from mainland China in 2005 (C. M. Chen and Tsai 2019, 84).

⁶ On a visit to the Beishan Broadcasting Wall in November 2022, a recording played one of Teng’s famous songs, “Honey Sweet” (甜蜜蜜), as well as this recorded message: “My dear comrades in mainland China, how are you? I am Teresa Teng. I am here at the broadcasting station on Kinmen Island to speak with you. What I want to say today

The Great Firewall

The Beishan Broadcasting Wall plays a key symbolic role in Malaysian artist Namewee's (黃明志) 2021 music video "The Wall" (牆外), which exceeded one million views within 18 hours after it was released on YouTube (Zhe 2021).⁷ The song and video offer a clear metaphor for restrictions on cross-strait communication. The music video is set in Kinmen and depicts the story of a budding relationship between two young children whose parents and life circumstances aim to prevent them from becoming friends. The girl, dressed in red to represent mainland China, is trapped in her home both by a wall surrounding her house and by her strict parents. Her part is sung by a mainland Chinese singer under the pseudonym Xiaohua (小花). The mischievous boy, whose part is sung by Namewee, dresses in blue to symbolize Taiwan as he navigates life with an alcoholic father as his only family. After meeting through a chance encounter, the children maintain their friendship despite the obstacles by tossing items over a wall that surrounds the girl's house: first the boy's misplaced book bag, then food, drawings, and medicine. The barrier created by this wall is broken not only through these material items, but also through sound, as indicated in the lyrics:

I'm waiting outside the wall for you
 I'm waiting for your reply
 I hear your breath and your heartbeat
 我站在牆外等妳
 期盼著你的回應
 我聽見你的心跳還有你的呼吸

The wall here creates an obstacle in the friendship of the two children, but it cannot fully prevent their exchanges. The sounds of their breaths and heartbeats continue to pass through the wall as they listen on opposite sides for signs of presence and aliveness.⁸

Given the political metaphor in this song, it is perhaps unsurprising that Namewee's music is censored in mainland China's highly regulated media environment. Indeed, the wall in the song is a clear parallel to mainland China's Great Firewall, or the current system of censorship and surveillance that the CCP uses to regulate the internet. This extensive system restricts access to foreign websites and filters or blocks content deemed politically sensitive or inappropriate by the Chinese authorities. Though far-reaching, sociologist Margaret Roberts explains that the Great Firewall is "porous" (Roberts 2017). This means that the mainland Chinese government makes censored information more difficult to access for the vast majority of

is that I am delighted to stand on the front line of my free homeland, Kinmen Island; I feel very happy and fortunate. I expect that our fellow countrymen in mainland China can also enjoy democracy and liberty just like us" (translation by author). Chen-ching Cheng writes that this message was originally broadcast at the Mashan observatory in the 1990s, not at the Beishan station (C. C. Cheng 2019).

⁷ Namewee. "There is an invisible WALL between us..Namewee 黃明志金門觀光主題曲 【The Wall 牆外】 Ft.小花 Flower@鬼才做音樂 2021Ghosician." November 12, 2021. <https://www.youtube.com/watch?v=WCLIFA9SiDI>. The song title's literal translation is "Outside the Wall."

⁸ This story of forbidden love evokes Shifu Wang's classic Chinese tale *The Story of the Western Wing* (西廂記), where in one scene two lovers forbidden from seeing each other communicate by reading love letters through a wall (Shifu 1991).

citizens by slowing down certain webpages, for example, or pushing politically controversial search results to the bottom of the page. Only a small minority of citizens who are tech savvy, wealthy, educated, and politically motivated are likely to expend the extra effort that is required to access censored information. By allowing this minority to bypass the firewall, mainland China's government avoids calling attention to the severity of the censorship (which could cause backlash), while also continuing to control the information that is accessible to most people. For those motivated to bypass the Great Firewall, a common strategy is to use a Virtual Private Network (VPN) to hide the physical location of their computer's server. This practice is known as "climbing the wall" (翻牆) (Ng 2013; Yu 2021).

Though Namewee's "The Wall" is banned on the mainland Chinese internet, people in mainland China still managed to climb the firewall and respond to the music video online. Some commenters said the song brought them to tears. One message read, "Listening to this from inside the wall, I am moved and also very sad. It is fate that has arranged for us to be born in this closed-off country, but longing for freedom and beauty is human nature. I am grateful that you understand, empathize, and offer comfort. Thank you for your understanding!" (墙内的我听了感动·也很难过·是命运安排我们出生在这个封闭的国度·但是向往自由和美好是所有人类的天性·感恩你能懂能体谅还给安慰·谢谢你明志!). Another comment said, "There is a very touching scene. The background is a wall full of holes, and Namewee sings through the wall. I hope that one day we can all break down the wall" (有一个镜头好带感·背景是一面全是洞的墙·明志把墙唱穿了"·希望有一天能破墙成功) (Zhe 2021; *World Journal* 2021).

At the end of "The Wall," the budding friendship between the two children is thwarted when the girl's family moves away, boarding boat headed toward mainland China. The boy climbs a wall and waves his blue jacket toward the boat, attempting to say a final goodbye. Shots of the boy waving are interspersed with Namewee performing in front and on top of the Beishan Broadcasting Wall, as he and the boy continue their persistent attempts to send messages across the Taiwan Strait despite restrictions.



Figure 2. Namewee in front of the Beishan Broadcasting Wall, from the music video “The Wall.” Screenshot from <https://www.youtube.com/watch?v=WCLIFA9SiDI>.

Sounding and Listening

Still standing tall today, the Beishan Broadcasting Wall is a symbol of the continuous nature of cross-strait communication, even paralleling contemporary technologies such as the Great Firewall. The loudspeakers housed in the holes (or “pores”) of this sound wall transform this formidable structure into a vehicle for communication. Similarly, while the Great Firewall is designed to block and control information, its openings present a site for exchange.

This wall of sound also represents the centrality of sound – of speaking and listening – in the restricted communication environment of the Taiwan Strait. Walls can create physical and visual partitions. However (as anyone who has ever had noisy neighbors can attest), walls have limited ability to prevent sound from traveling through. Foregrounding sound as a way of understanding cross-strait relations then invites a reimagining of space, distance, regulation, and restriction, as sound often travels through physical barriers in ways that other communication mediums cannot.⁹

Throughout 75 years of cross-strait relations, communication between people in Taiwan and mainland China has often centered around sound. Even amidst the highly regulated media environments of the 1950s and 1960s, loudspeaker and radio broadcasts projected voices across the strait, encompassing official propaganda as well as informal exchanges. By the late 1990s, there was a shift away from loudspeaker and radio broadcasts and toward new modes of internet

⁹ There is abundant literature discussing the relation of sound and space. Much of this centers on how sound reconfigures perceptions of public and private space (as with the example of the noisy neighbor). See for example Abe 2018 and Born 2013. For more general discussions of sound and space, start with Lefebvre 1992; Connor 1997, 206; Feld and Basso 1996.

communication. Even with these new possibilities, sound and voice communication continued to present unique possibilities for people in Taiwan and mainland China to speak and listen directly to one another.

Because of the persistence of sound in crossing the Taiwan Strait, listening is a central approach for this research. I focus specifically on instances where people's voices circumvent restrictions, proposing a counternarrative to the emphasis on political tension in the Taiwan Strait by emphasizing the on-the-ground experiences of individuals. Listening as an analytical approach has long been used to draw out new ways of understanding the world, frequently challenging mainstream narratives based on visual perceptions. The edited volume *Hearing Cultures* is representative of the emergence of a body of thought, often referred to as sound studies, which aimed to bring attention to sound as a way of knowing and understanding the world (Erlmann 2004).¹⁰ This volume began with the question, asked originally by anthropologist James Clifford in 1986 and restated by musicologist Veit Erlmann, "But what of the ethnographic ear?" (Erlmann 2004, 1; Clifford 1986, 12). This question pointed toward possibilities for conceptualizing "new ways of knowing a culture and of gaining a deepened understanding of how the members of a society know each other" (Erlmann 2004, 3). The realization that prioritizing listening and sound could provide a new analytical perspective unlocked a range of new frameworks and methods for understanding how people relate to one another through sound.

Taiwan offers an intriguing environment from which to listen. During my initial visit to Taipei in 2018, I was consistently struck by the extent to which the soundscape differed from what I was used to in my daily life in Berkeley. The subway trains and trash trucks came with their own little jingles, the tea kettles and doorbells sang as well, and the abundant motorcycles roared to life at the change of a red light.¹¹ On my second trip to Taiwan in 2019, this time for intensive language study and armed with two years of graduate coursework, I listened not only with "ethnographic ears" but also with "postcolonial ears," defined by ethnomusicologist Thomas Solomon as how "power relations are *sonically* constructed, for example how representations of difference are embodied in sound" (Solomon 2012, 235). While visiting my Taiwanese American partner's grandparents' house in Taipei, I began to understand the mix of languages spoken there (Taiwanese Hokkien, Mandarin, Japanese) in the context of Taiwan's complex and overlapping colonial histories. I listened to the four languages included in public transit announcements as a contemporary reminder of how histories of power continuously influence the present.¹²

By my third trip to Taiwan for a year of dissertation research (April 2022-March 2023), I began to listen not only to the sounds of streets, stores, and homes, but also to utilize my improved Mandarin language skills to listen to people's voices. I listened to how reporters, taxi drivers, and friends discussed current events and the ways in which politics influenced their daily lives. I listened not only to the content of what people said, but also to how they spoke: their tone, pauses, vocables, code-switching, and other affective expressions conveyed through their voices.

¹⁰ Anthropologist Steven Feld coined this concept "acoustemology" (Feld 2017).

¹¹ For discussions of Taipei's soundscapes, see Guy 2021, 2019; Hsieh 2019, 2017, 2021.

¹² Taiwan's colonial history includes Dutch colonization in the seventeenth century and the Japanese colonial period (1895-1945). As suggested by Ping-hui Liao, the KMT can also be considered a colonial government. He writes, "The [indigenous people] in Taiwan have been victims of different periods of foreign and domestic colonial cultures: Dutch, Spanish, southern Chinese, Japanese and KMT" (Liao 1999, 200).

I was interested specifically in how these voices aligned with or differed from the perspectives represented in media coverage of cross-strait relations. What would I find by listening beyond narratives emphasizing only conflict and political tension?

What I discovered was a continuous series of interactions in which people in Taiwan and mainland China use various forms of audio communication technologies to speak and listen to one another. From this broad listening perspective, I zoomed in to specific case studies and the platforms that hosted these interactions. These mediums range from loudspeakers and radios beginning in the 1950s to social media and videogames today. Each of these audio communication mediums creates different affordances for cross-strait interactions, as well as different limitations. Drawing on Andrew Jones' concept of "circuit listening," I consider how the infrastructures behind these audio communication technologies inform the conditions of possibility for cross-strait interactions as they "operate across, within, and in complex and often complicit relationships with existing social topologies" (2020, 7).¹³

The idea of "circuit listening" further draws attention to Taiwan's contributions to the production of audio communication technologies. Taiwan has long played a central role in the global hardware production industry, tracing back to its development into a manufacturing hub for transistor electronics in the 1960s (Rigger 2021, 69). In recent decades, Taiwan has established itself as an indispensable player in semiconductor manufacturing, as exemplified by industry leader Taiwan Semiconductor Manufacturing Company (TSMC). That Taiwan alone is capable of producing the world's most advanced semiconductor microchips, which in turn power the most advanced computer technologies (including an exclusive partnership with Apple), has led this industry to gain a reputation as Taiwan's "silicon shield." The potential economic repercussions of a disruption to this industry are then discussed as great enough to deter aggressive actions from mainland China (Cronin 2022; C. Miller 2022). In the context of cross-strait communication, tuning in to Taiwan's longstanding leading role in the development and production of computer hardware highlights the centrality of Taiwan in producing the technologies that both enable and prevent cross-strait communication, and points toward how these technologies are deeply intertwined with its current political situation.

Chapter Overview

Each of the following chapters examines cross-strait audio communication from different historical and technological vantage points. Following this introduction and its brief discussion of cross-strait loudspeaker broadcasts, Chapter 2 focuses on relationships formed through and surrounding radio broadcasts during a transitional period in cross-strait relations, from the late 1980s through the early 2000s. During this period, mainland China underwent rapid economic reform amidst growing activist movements, and Taiwan shifted toward democratic governance and a more open media environment. Cross-strait relations opened to a degree not seen in decades. People in mainland China and Taiwan could now call each other on the phone or even visit in person, and letters were more likely to arrive at their intended destinations. I explore these shifts in the context of the Taiwanese radio program Voice of Asia (亞洲之聲), which had a dedicated following of mostly teenaged listeners in mainland China for its culture-focused content and coverage of Mandarin-language popular music from Taiwan. Increased opportunities

¹³ Regarding "audible" infrastructures, see also Devine and Boudreault-Fournier 2021.

for exchange surrounding these broadcasts in turn led to sustained listener-host relationships lasting many years.

Broadcast radio is often conceptualized as a unidirectional flow of information: a host sends information out to listeners without necessarily expecting a response. In this chapter I seek to expand the definition of broadcast radio to include not only the broadcasts, but also listeners' responses to these broadcasts through letters, phone calls, and visits. I conceptualize broadcast radio as a multimedia, multidirectional assemblage in which both broadcasters and listeners play an active role. With a focus on Voice of Asia, I demonstrate how these exchanges contributed to shifts in what I call the "cross-strait socio-political imaginary," or how broadcasters and listeners on opposite sides of the Taiwan Strait conceptualized and imagined each other. I also point toward how these interactions established a model for future exchanges on different mediums, especially as radio use declined in the late 1990s and opportunities for communication via the internet became more widespread.

In Chapter 3, I fast forward to February 2021 to discuss one such instance of internet mediated cross-strait communication. I focus on the "Clubhouse moment," when internet users in mainland China, Taiwan, and across the world flooded onto the audio-centered social media app Clubhouse for a period of only a few weeks. On Clubhouse, people joined audio chatrooms with as few as two or as many as 5,000 participants, with topics ranging from current events to learning languages to sharing opinions on cross-strait relations. This brief moment of open communication was significant in the context of declining cross-strait relations since 2015, which materialized in large declines in the number of mainland Chinese tourists and exchange students in Taiwan, in turn leading to fewer opportunities for exchange.

Though short-lived, I examine how the virality of the Clubhouse moment served as a rare opportunity for internet users in Taiwan and mainland China to speak and listen directly to one another before mainland China's censors caught on and banned the app. Since the platform centers around live voice communication, Clubhouse users often discussed increased engagement, empathy, and emotional reactions while listening to others on Clubhouse, even if the ideas being shared contrasted with their own beliefs. While social media often consists of "echo chambers" in which internet users are exposed only to ideas similar to their own, in this chapter I demonstrate how the Clubhouse moment presented a brief contrast to this environment. Clubhouse users, I contend, were able to speak and listen to people whose ideas differed from their own, leading to the brief formation of a cross-strait public sphere in which diverse voices and opinions were welcome.

In Chapter 4, I shift focus from the previous exploration of how people connect through voice communication to a detailed analysis of moments marked by difference. I explore these dynamics in the context of live voice communication in online videogames. While the Clubhouse moment was short-lived, online videogames serve as a smaller scale but longer lasting opportunity for cross-strait communication as gamers in mainland China take advantage of gaps in the Great Firewall to connect with gamers abroad. While videogame interactions between Taiwanese and mainland Chinese gamers are often hostile (as has often been the case with the videogame World of Warcraft), in this chapter I instead focus on exceptions to this hostility in which Taiwanese and mainland Chinese gamers find the opportunity for in-depth, live voice exchanges within the relaxed environment of online play.

In some cases, these interactions are livestreamed, recorded, and then posted for a public audience on video streaming sites such as YouTube and Bilibili. Through close analysis of these

livestreamed recordings, I demonstrate that in contrast to associations of videogames with violence, a relaxed environment of online play can in fact lead to increased engagement and even empathetic exchanges. These informal, relaxed interactions then provide a site for gamers to confront cultural, linguistic, and political differences. Specific to language, I highlight a trend in which mainland Chinese gamers often perceive Taiwanese Mandarin accents to be “cute.” At the same time, Taiwanese gamers often disagree with this characterization. By delving into these differing perceptions of languages and accents, I draw out multiple layers of complex power relations as understood through live-voice communication.

The final chapter develops the theme of regulation, a common thread throughout the dissertation. I specifically focus on the repeating pattern of “regulatory latency” in which communication technologies often develop before their corresponding regulating technologies. For example, radio broadcasts (discussed in Chapter 2) served as a site for cross-strait exchange until the late 1990s, when regulatory mechanisms in mainland China began to more comprehensively block radio broadcasts from Taiwan. At the same time, the spread of the internet introduced new opportunities for unregulated interaction. In a similar pattern, the social media app Clubhouse (Chapter 3) quickly became popular as a site for cross-strait exchange, allowing a brief window of open communication before mainland China’s government implemented corresponding regulation and banned the app.

In analyzing regulatory policies and technologies in the contrasting media environments of Taiwan and mainland China, I show how regulation is not in and of itself positive or negative, and instead is a tool that serves different purposes in different contexts. In mainland China, media regulation is synonymous with censorship and surveillance, creating an environment in which government ideologies are prioritized at the expense of open communication and freedom of expression. In contrast, Taiwan since the 1980s has maintained a relatively open media environment. In this context, regulation serves the purpose of combatting disinformation to maintain spaces for democratic discourse. While regulatory latency in Taiwan’s media environment can lead to increased spread of disinformation, in mainland China delayed regulation can in fact open a brief window for open communication. This chapter then continues to analyze this pattern in the context of currently developing communication technologies, which increasingly involve artificial intelligence (AI) tools. I suggest that the development of audio AI tools in the past two decades has often lagged behind similar tools for text and images. This trend then leads to longer periods of latency in both the development of audio technologies and regulation of these technologies. This association of the development of audio technologies with latency, I argue, reveals that audio communication in the age of AI may be especially well suited to communication outside of the bounds of regulation, leading to longer periods in which people in highly regulated media environments can engage in open communication.

Research Methods

The research for this dissertation was conducted from fall 2021 through spring 2023, including a year-long period in Taipei from April 2022 through March 2023. Due to visa delays caused by the Covid-19 pandemic, the initial six months of research (September-March 2021) took place in Berkeley. During this initial period, I took advantage of local resources, in part by connecting with professors in Berkeley’s computer science department to gather background

knowledge on developing audio AI technologies. I also conducted interviews with exchange students from mainland China and Taiwan, as well as remote interviews with individuals in Taiwan. I carried out preliminary exploratory research during trips to Taiwan in the spring of 2018 and summer of 2019. In preparation for this research, I also participated in intensive Mandarin language study from 2018 through 2022, both at UC Berkeley and the International Chinese Language Program at National Taiwan University.

Each chapter of this dissertation required distinct methods for collecting and analyzing research materials. The primary method was to conduct interviews, mainly with Taiwanese and mainland Chinese internet users with experience on Clubhouse and/or videogames, as well as with professionals in the field of artificial intelligence. In Taiwan, I met interlocutors for interviews either at their workplaces or at cafes. Chapter 2 involved archival work at Radio Taiwan International in which I studied listeners' letters and program transcripts, placing these materials in dialogue with interview accounts from key interlocutors. Chapters 3 and 4 incorporated online observation on Clubhouse, Youtube, and Bilibili, often alongside and in dialogue with mainland Chinese and Taiwanese participants.

Following the guidelines of UC Berkeley's Institutional Review Board (IRB) and to ensure the safety of interlocutors, most people interviewed for this research remain anonymous. Exceptions were made for public figures (journalists and some AI researchers) who are easily identifiable and gave written consent for the use of their names. Anonymous interlocutors are referred to by coded initials unrelated to their actual names.

To analyze collected materials, I transcribed and annotated interviews and connected these transcripts with observational and archival materials. For interview recordings, I noted not only linguistic meaning but also details such as tone of voice, pauses, and code-switching. I also drew on specific methodological listening techniques. One such framework was affective listening, which proved especially relevant in the context of Clubhouse (Chapter 3). Affective listening involves engaging with the emotional and relational dimensions of auditory experiences (J. A. Wang 2016). Clubhouse users often experienced this phenomenon, discussing deep emotional reactions and connections in part facilitated by the app's emphasis on live voice communication. Affective listening was also relevant in the radio interactions discussed in Chapter 2, where listeners and radio hosts formed long-lasting connections through multimedia communication surrounding radio broadcasts. In the context of live voice communication through videogames (Chapter 5), affective listening provided an analytical approach to understand interactions in which gamers from Taiwan and mainland China confront their differences based on the sonic expressions of each other's voices.

I also drew on the methodological approach of "listening to listening," developed here especially in the context of videogame interactions (Eidsheim 2019; Feld 2015). This analysis involved not only analyzing the content of recorded livestreams, but also tuning in to how Taiwanese and mainland Chinese gamers perceived and reacted to this content. By listening to videogame interactions alongside Taiwanese and mainland Chinese gamers, I developed a sense of how gamers related to one another, as well as how audiences of these recorded livestreams interpreted these interactions. This method was similarly important in the radio and Clubhouse chapters (Chapters 2 and 4), where interviews combined with observational and archival research supplied layered and intersecting accounts of individuals' experiences. Listening to listening served as a way to begin to break down the political processes embedded in the voice and to closely examine how people grapple with political differences in dialogue with others.

Musicologist Nina Sun Eidsheim writes, “Because listening is never neutral, but rather always actively produces meaning, it is a political act. Through listening, we name and define” (Eidsheim 2019, 24; Owens 2023). Listening across the Taiwan Strait is then an attempt to name, define, and explain the complex processes through which politics intersect with everyday life as people in Taiwan and mainland China persistently articulate, project, and confront ongoing experiences with political tension and ideological difference.

2. The Golden Era of Cross-Strait Exchange: Multidirectional Broadcast Radio from the Late 1980s through Early 2000s

The late 1980s marked a pivotal time in cross-strait relations, characterized by significant socio-political changes and more opportunities for exchange. The end of martial law in Taiwan in 1987 was paired with unprecedented freedoms, including the ability to travel and make phone calls across the Taiwan Strait. In mainland China, this period was marked by economic development and growing activist movements, culminating in the tragic events of Tiananmen Square in 1989. Though the event led to a brief pause in cross-strait communication, it also strengthened a sense of solidarity between people on both sides of the strait.

The effects of this transformative period were acutely felt on Taiwan's radio broadcasts, which not only reported on these increased opportunities for cross-strait exchange, but also created the very conditions for these interactions. This chapter delves into this transformative era through a focus on Taiwan's government-owned national broadcasting station, Radio Taiwan International (RTI, 中央廣播電台), and specifically its culture-focused station Voice of Asia (亞洲之聲), during what has been called the "golden era" from the late 1980s through the early 2000s.¹⁴ Taking the golden era of RTI as a case study illustrates the shifts in cross-strait relations during the late 1980s and 1990s, from a period of closed-off communication to one of greater openness. As I show in this dissertation, cross-strait dialogue among individuals has persisted throughout decades of shifting geopolitics, in conjunction with developments in communication technologies, and amidst fluctuating periods of restriction and openness.

While broadcast radio is often conceptualized as a one-directional medium that sends information out into the world without necessarily receiving information in return, here I consider broadcast radio as a multimedia, multidirectional assemblage that includes both the broadcasts and listeners' responses to these broadcasts via letter, phone call, and in-person visits. Direct exchanges through and surrounding RTI's broadcasts contributed to shifts in the cross-strait socio-political imaginary, influencing people's perceptions of their neighbors across the strait and contributing to how they imagined their own beliefs in relation to others. These exchanges were notable not only for their content, but also for their novelty; the very act of communication often felt as significant as the information being communicated. These multimedia exchanges also established a model for ongoing interactions that would continue into future decades and through the adoption of new communication technologies.

¹⁴ Radio Taiwan International has gone through various name changes throughout its history. It was founded as the Central Broadcasting Station (CBS, 中央廣播電台) in 1928. In 1939, Voice of China was established under CBS and began international broadcasts. In 1946, the Broadcasting Corporation of China (BCC, 中國廣播公司) was established, and Voice of Free China was created under the BCC's international department. After the KMT moved to Taiwan, CBS continued to broadcast, now directed toward mainland China and housed under the Ministry of Defense. The Voice of Asia was inaugurated in 1979, also under the BCC's international department. From 1996-1998, the international department of BCC, including Voice of Free China and Voice of Asia, merged with CBS. Voice of Free China was renamed Radio Taipei International. In 2002, Voice of Asia ended, and Radio Taipei International was renamed Radio Taiwan International. Despite these changes, the Chinese name of Radio Taiwan International (中央廣播電台) still translates to Central Broadcasting Station (Rawnsley 2000, 124-30; L. Li and Gregory 2009). Due to this complexity, in this chapter I refer to what were formerly known as CBS and BCC by the current English name, RTI.

The analysis in this chapter is based on research in RTI's archive (February-March 2023), during which I listened to recorded Voice of Asia broadcasts and sifted through letters from listeners across three decades (1980s-2000s). This research is also informed by interviews with two key interlocutors. The first is retired Voice of Asia broadcaster Wu Rui-wen (吳瑞文) who started his radio career in 1984 and continued for 30 years, also serving for a period as director of RTI. The second is Wu Rui-wen's long-time listener and friend Cindy, who listened to Voice of Asia broadcasts in mainland China beginning in 1987, when she was only 10 years old. At that time, mainland China's border and travel restrictions were beginning to loosen, and her father left the country and brought back a transistor radio. Cindy remembers that there were not many broadcasts in mainland China at the time, so she was attracted to Taiwanese shows and especially liked the popular music programs. As an adult, Cindy moved to Taiwan and later reconnected with Wu Rui-wen.¹⁵ With this combination of materials, I consider not only how official policies regulating cross-strait interaction shifted during RTI's golden era, but also how these policies played out on the ground as Taiwanese and mainland Chinese people's perceptions of each other were influenced by new opportunities for direct communication.



Figure 3. Radio Taiwan International in Taipei, Taiwan. Photo by author.

¹⁵ Interview with author, March 10 2023.

Radio Before the Golden Era

The golden era of RTI stands in contrast to the prior period of cross-strait radio beginning in the 1950s, during which centralized, highly regulated broadcasts served primarily as tools for psychological warfare. The transition to more culturally focused media in the 1970s marked the beginning of a broader shift toward deregulation and decentralization of both the media and government on both sides of the Taiwan Strait, paving the way for increased opportunities for interaction by the late 1980s.

While the use of radio for wartime purposes may seem striking today, radio was in fact invented largely as a wartime technology.¹⁶ Indeed, RTI was established by the Kuomintang (KMT) nationalist government under leader Chiang Kai-shek, which at the time governed the Republic of China (ROC) in mainland China and was headquartered in Nanjing.¹⁷ It was originally known as the Central Broadcasting Station (CBS). Though radio was first used by the KMT for internal military communication and to disseminate information locally, the development of shortwave technology in the 1930s meant that information could travel longer distances. With this new shortwave technology, RTI founded the station Voice of China and began to broadcast within China and to the outside world (L. Li and Gregory 2009; Rawnsley 2000, 124–25).

During and after World War II, RTI followed the KMT from city to city, finally settling in Taipei in 1949 following the KMT's retreat in the Chinese Civil War. Upon establishing RTI in Taiwan, the KMT government took over the broadcasting infrastructure left by the Japanese colonial administration that had occupied Taiwan for 50 years (1895-1945).¹⁸ Ironically, this infrastructure had been used by the Japanese colonial government during World War II in efforts against the KMT in mainland China (L. Li and Gregory 2009).

In Taiwan, the KMT's radio efforts were split between psychological warfare directed toward mainland China and news or cultural programming directed toward international and overseas Chinese audiences, now under the name Voice of Free China (自由中國之聲) (Rawnsley 2000, 125). There was overlap in audience between these efforts, with listeners in mainland China frequently tuning in to Voice of Free China programs. In this context, shortwave radio broadcasts significantly extended the reach of psychological warfare. While loudspeaker “walls of sound” in southern China and the KMT-occupied islands in the Taiwan Strait were limited in

¹⁶ International radio broadcasts were largely initiated to meet wartime aims. Political scientist Gary Rawnsley details how international radio broadcast emerged as early as 1926 and expanded during World War II, though it was during the Cold War that radio made its “largest contribution to the dissemination of propaganda” (Rawnsley 1996, 8). In the context of China, historian Paulina Hartono explains that the use of radio was at first highly regulated to be used only for wartime purposes: “The development of radio in China had a long, close relationship with war. Because radio technology was used by the military to communicate tactical movements, or to monitor enemy transmissions, it was classified as war contraband by the Nationalists as early as 1923, a move that largely limited civilian purchases for the next decade” (Hartono 2020, 29).

¹⁷ The Republic of China (ROC) was established in 1912 by the KMT nationalist government in mainland China. When the KMT retreated to Taiwan after the Chinese Civil War, they established the ROC in Taiwan.

¹⁸ CBS's most powerful shortwave broadcasting station in Taiwan was in Minxiong, Chiayi County, and had been set up by the Japanese colonial government in 1938 “in order to keep Taiwanese from listening to CBS broadcasts from Nanjing and to wage psychological warfare against mainland Chinese” (L. Li and Gregory 2009).

range, radio broadcasts reached far beyond the Taiwan Strait, amplifying the KMT's influence and efforts during this period.¹⁹

While the KMT nationalist government developed RTI in the 1930s, the Chinese Communist Party also invested in radio as a tool for internal communication. During the Chinese Civil War, radio stations on opposing sides relied on different strategies to appeal to their populace, with the CCP emphasizing emotional identification “to draw an unfavorable contrast between the emotionally restrained Chiang Kai-shek and Mao Zedong” (Hartono 2020, 30; Perry 2013, 112). After the Chinese Civil War, radio continued to be used in this way, now mobilized on a larger scale to reach the masses. In spreading the official state line to the far reaches of the country, communist radio became “a crucial instrument in establishing and shaping the relationship between the authority at the top and the general public at the bottom” (Lei 2019, 51). While the KMT took over broadcasting infrastructure from Japan upon arriving in Taiwan in 1949, the CCP in mainland China took over radio stations from the KMT, aiming to expand the reach of radio beyond the urban elite. The CCP developed an extensive wired broadcasting network, with “ubiquitous loudspeakers that revolutionized the urban and rural soundscape, redefining time and space, political mobilization, and everyday life” (Jie Li 2020, 26). This infrastructure was built for collective listening, minimizing the need for every household to own a radio while also granting the government control over the content of broadcasts. Mainland China's wired broadcasting and loudspeakers were therefore used internally “to construct an unprecedentedly pervasive socialist media network,” providing a way for the government to communicate with its citizens while largely maintaining media isolation from the rest of the world (Jones 2020, 7).

This era of ideological broadcasting was not limited to the Taiwan Strait, and in fact was representative of broader Cold War trends. During this period, major international broadcasters like Voice of America and the British Broadcasting Corporation (BBC) aimed to promote Western democratic values and counter Soviet influence, creating a global network of ideological dissemination. This period saw radio transform into a powerful medium of psychological warfare, where the battle for hearts and minds was fought over the airwaves, shaping public opinion and political landscapes around the world (Rawnsley 1996, 12–17).

The representation of competing political views on the airwaves created opportunities for people to access diverse perspectives despite governments' attempts to control the media of their citizens. In mainland China, even while the CCP maintained strict control over the media environment, people managed to procure portable transistor radios and illegally tune in to what was called “enemy radio” (敵台) from abroad. Although this practice began in the 1950s and 60s, it became especially widespread in the 1970s with the increased ubiquity of these portable devices.²⁰ This illegal practice then led to more opportunities for people to access information outside of mainland China's wired broadcasting network by tuning in to shortwave broadcasts of programs such as Voice of America, the BBC, and Voice of Free China. As Chinese literature

¹⁹ See Chapter 1 for a more detailed discussion of loudspeaker “walls of sound.”

²⁰ Broadcasts from Taiwan were accessible in mainland China as early as 1955. Yu Wang writes, “In the early 1950s, radio stations in Taiwan had very limited reception in the mainland. This situation changed in 1955 when the nationalist government erected a medium wave launch station with an output power of 600 kilowatts per hour. Its signals could cover all of China as well as Southeast Asia” (Y. Wang 2019, 152).

scholar Jie Li explains, while wired radio and loudspeakers worked to amplify the “throat and tongue of the state,” transistor radios worked to “extend the ears of the populace” (2020, 44).

The ability of mainland Chinese listeners to tune in to enemy radio from Taiwan facilitated cross-strait interactions between listeners and broadcasters (albeit on a small scale), most often in the form of letters. Especially popular in the 1960s was a program called Listeners’ Mailbox (聽眾信箱). Through this program, mainland Chinese listeners could write letters to the station, often asking for assistance in locating missing relatives or friends. Historian Yu Wang explains that Voice of Free China originally set up a mailbox in Taipei and received the first letter two weeks after its first broadcast. Encouraged by this response, they then set up mailboxes in Bangkok and Manila, both as a strategic move to protect letter writers from the authorities and as a way to help the station to receive more of the letters. Still, many of the letters were lost or confiscated. In one case, a radio listener wrote 13 letters but only one arrived at the designated address, with the rest commandeered by the police. This means that while Voice of Free China *received* 509 letters in 1962, the number of letters *sent* was likely much higher (Y. Wang 2019, 154–55). This example indicates the high degree to which cross-strait communication was regulated during the 1950s and 1960s, while also revealing that even in this highly regulated environment there were messages that managed to slip through.

While confiscating letters was one form of preventing cross-strait interactions, the CCP also took considerable efforts to directly regulate listeners’ access to radio broadcasts. These efforts took two forms: controlling the sale and use of radio sets and jamming radio signals. Jie Li writes that in 1950 the CCP monitored the sale and registration of radio sets while “mobilizing 800 radio technicians to physically disable the shortwave function on all the radio sets they could get their hands on” (2020, 29). The CCP also tried to block programs arriving from outside of its borders through the practice of jamming, or intentionally transmitting radio signals on the same frequencies as these international broadcasts to disrupt their reception and reduce the content to static. That said, the output power of the jamming stations in mainland China paled in comparison to that of the stations they aimed to block.²¹ International stations would also regularly change their broadcasting frequencies, initiating a game of cat and mouse. While these methods may have deterred some listeners, accounts and letters at the time indicate that listening to Taiwanese radio was still widespread. This is evidenced not only by the letters received by Voice of Free China, but also by reports of people listening to Voice of Free China in mainland Chinese newspapers.²²

²¹ Yu Wang writes, “The output power of these jamming stations was considerably lower than that of China’s rivals, such as the 600-kilowatt medium-wave launching station in Taiwan or the 1000-kilowatt medium-wave and 35-kilowatt short-wave launching stations in Okinawa, Japan. The 18 medium-wave launching machines in Tianjin had an average output power of 530 watts. As a result, they were only able to jam around 6 of the 20 frequencies from the Voice of America. Even worse were Tianjin’s two short-wave stations, which proved incapable of blocking any signals at all” (2019, 152–53).

²² Yu Wang explains, “Voice of Free China’s reach in mainland China particularly demonstrated the consequences of this technological gap. Its mainland listeners often referenced it when publicly talking about and criticizing current social and economic problems. In June 1955, one influential newspaper noted that many people listened to the Voice of Free China to enjoy “yellow music,” a term used to refer to erotic or indecent music pieces” (Y. Wang 2019, 153).

Former director of RTI Zhang Zheng (張正) explained in a blog post the dynamic of shortwave broadcasts and attempts to block them by invoking the lyrics of a 2010 love song by Taiwanese singer Jonathan Lee (李宗盛), entitled “A Song for Myself” (給自己的歌):

It can't be forgiven, it can't be blocked, hatred climbs over the wall at night
 It's completely empty, but it buzzes with sound, who put a cold weapon in your heart?
 [是不能原諒·卻無法阻擋·恨意在夜裡翻牆
 是空空蕩蕩·卻嗡嗡作響·誰在你心裡放冷槍?]

Though this song is about lingering hatred for a past lover, Zhang shows that the theme is unexpectedly consistent with attempts by mainland Chinese authorities to block Taiwan's shortwave broadcasts. Since there is no way to fully block these broadcasts, sounds from Taiwan can climb the wall, “buzzing with sound.” The metaphor of “climbing the wall” (翻牆) in this song creates another layer of complexity. This metaphor is most often used in the context of mainland China's controls over the internet to indicate various methods people use to illegally access internet from abroad. This same metaphor invoked in the context of shortwave radio suggests that listening to enemy radio established an earlier precedent for bypassing media restrictions. Parallels to the song go even further in that it invokes “climbing over the wall at night,” since shortwave signals tend to be clearer and travel farther at night (Z. Zheng 2020). This analogy then indicates that the technologies and resources to regulate enemy radio broadcasts could not keep up with the transmission power of the broadcasts themselves, thus opening a window for mainland Chinese listeners to access radio from abroad.

The practice of listening to enemy radio was also present in Taiwan, albeit on a smaller scale. In the first decade under Chiang's rule in Taiwan, people were prohibited from accessing unauthorized news, especially from mainland Chinese newspapers or radio broadcasts, and could not receive mail from relatives living in mainland China (Chai 2000, 134). Radio in Taiwan was thus highly regulated, though full regulation of this information was impossible. As historian Dayton Lekner demonstrates in the context of loudspeaker broadcasts, on the ROC island of Kinmen people would often hear loudspeaker broadcasts drifting across the strait, and they were simply instructed not to actively listen to or discuss these broadcasts (2022, 14–15).

By the 1970s, transistor radios became more widespread in mainland China and Taiwan. At the same time, programming on enemy radio began to soften in tone, in line with political developments at the time (Jones 2020, 176). In 1971, Beijing displaced Taipei as the representative of China in the United Nations. In 1979, the United States followed suit, normalizing diplomatic relations with the PRC and moving the US Embassy from Taipei to Beijing. While international recognition of Beijing in lieu of Taipei was a blow to Taiwan's international status, the easing of tensions between the US and mainland China also led to a gradual easing of tensions across the Taiwan Strait (Yeh 2013, 6).

In line with a shift to culture-focused media, 1979 also marked RTI's inauguration of Voice of Asia. Voice of Asia host Wu Rui-wen explained that while Voice of Free China had existed for decades to provide news and political information to international listeners, Voice of Asia was founded with a focus on popular culture. Voice of Asia was originally intended for a Southeast Asian and overseas Chinese audience. They weren't expecting to attract a mainland

Chinese audience, and yet their programs soon gained a large following in mainland China, as evidenced by the large volume of letters they began to receive from listeners.²³

The founding of Voice of Asia and its cultural focus merely solidified the role of cultural programming in Taiwan's broadcasts, since at the time of its founding this type of programming was already increasingly common. One of the main draws for mainland Chinese listeners in the 1970s was Mandarin-language popular music from Taiwan, and specifically the voice of Taiwanese singer Teresa Teng (鄧麗君). Accounts from mainland Chinese listeners discuss feeling intimate connections with Teng as they huddled around portable radios to listen to her songs. Teng's voice has been discussed as contrasting the "strident tones of the collective" that were so often heard in mainland China in the form of revolutionary songs transmitted over wired broadcast systems. Teng in contrast represented "sensuality, softness, and emotionality" (Jones 2020, 185). Chinese literature scholar Andrew Jones explains:

In nearly all accounts of encountering Teresa Teng's music for the first time, there is an extraordinary and almost overwhelming sense of affective intensity. Teng's soft pop stylings—which may register today as pleasantly romantic at best and perhaps syrupy or sentimental at worst—were perceived as nothing short of revelatory, a sound to be savored with a rush of tears (2020, 185).

That said, Jones complicates this understanding by connecting the novelty of Teng's voice with emerging technology at the time: the intimacy perceived in her voice was as much a contrast to socialist soundscape as it was a representation of the adoption of close-micing techniques and portable electronics. A microphone held next to Teng's mouth could be transmitted to a transistor radio held next to a listener's ear. The distance closed up, creating a sense of intimacy enabled as much by changing political situations as by new developments in audio technologies.²⁴

While political ideologies were transmitted implicitly through Teng's voice through its contrast to socialist aesthetics, cultural programming in the 1970s and 1980s still often came alongside explicit political messages. As a 1987 letter from a mainland Chinese listener attests:

I really like Teresa Teng's songs, but after they became popular in Shanghai and the CCP restricted them, some people secretly listened through Taiwan's radio broadcasts and recorded them themselves. While secretly listening, they also heard about the situation of Taiwan's Three People's Principles to unify China (Huang 2022, 58).²⁵
[很喜歡鄧麗君的歌曲·但中共限制在上海流行後·有人偷聽臺灣廣播·就自己錄·在偷聽歌曲過程中也聽到台灣搞三民主義 統一中國的情況。]

Teresa Teng and her popularity in mainland China can be thought to have set off a wave of Mandarin-language pop music coming out of Taiwan that gained a wide audience in mainland

²³ Interview with author, March 3 2023. Rawnsley writes that Voice of Asia received 420,000 letters from its audience over the course of 18 years (Rawnsley 2000, 128).

²⁴ For English-language discussions of Teresa Teng, see also (Schweig 2022; Cheng 2019; Wu 2021).

²⁵ The "Three People's Principles" refers to the three principles established by ROC founder Sun Yat-sen as the basis for his vision of Chinese society: nationalism, democracy, and social welfare. These three principles were central to Taiwan's then president Chiang Ching-Kuo's official policy to unify Taiwan and mainland China.

China in subsequent decades. While radio was no longer limited to explicit political ideology or information warfare, it still served as a medium to spread Taiwanese popular culture, and with it, Taiwanese values surrounding capitalism, gender roles, and the expression of emotion—a phenomenon anthropologist Marc Moskowitz calls Taiwan’s “musical counter invasion of China” (2009).

The intimacy and affective connections that Teresa Teng’s listeners discuss in relation to her voice also indicates a qualitative shift in how people related to radio broadcasts on an emotional level. Whereas radio was previously utilized and broadly understood as a tool for warfare, whether tactical or psychological, with the arrival of Teng’s voice in mainland China it also became a tool for the formation of intimate relationships. Teng’s voice on enemy radio set the stage for a new era in mediatized cross-strait relationships, providing a new platform from which people in mainland China and Taiwan could imagine each other and perceive their own beliefs in relation to one another. This trend would continue as opportunities for communication opened up through the 1980s and 1990s.

Contextualizing the Golden Era: Shifting Cross-Strait Socio-Political Imaginaries

While Voice of Asia gained a following in mainland China throughout the 1980s, it was not until 1987 with the end of martial law in Taiwan that the broadcasters realized the extent of their audience. This was because previously listeners’ letters from mainland China were sent through Hong Kong and then processed by the KMT Ministry of Culture upon landing in Taiwan. In the 1980s, the system changed: letters from mainland China were now forwarded from Hong Kong to be received directly by Voice of Asia broadcasters. Wu Rui-wen explained that the Voice of Asia broadcasters were surprised when they finally started to receive letters directly. He said, “It was only when we began to read so many letters from these listeners in mainland China that we realized, ‘Wow, it turns out so many people in so many places in mainland China are listening to radio programs from Taiwan’” [我們開始大量的閱讀了大陸的這些受眾的來信，才瞭解說，哇，原來在中國大陸那麼多地方，有那麼多的人在收聽來自於臺灣的廣播節目].²⁶

Since Voice of Asia measured their success in part on how many letters they received, the broadcasters were thrilled when they began to receive so many letters. Wu Rui-wen remembered:

In the two or three years before the 1989 democracy movement, we received an average of 1,000 to 2,000 letters from mainland China every month. I still remember that we once received 3,000 letters from mainland Chinese listeners in one month. We were so happy that we had a cake. We really couldn’t finish reading all of the letters.”²⁷

[八九民運之前的這兩三年呢，我們大概平均每一個月來自中國大陸的信件 1000 2000 封，我還記得我們曾經為了一個收到 3000 封大陸聽友的來信。呢我們切蛋糕，很開心。那信件簡直是都看不完啊]。

July 15, 1987, marked the end of 38 years of martial law in Taiwan, or the prolonged era of political repression referred to as the period of White Terror and characterized by political repression and human rights abuses under the authoritarian rule of the KMT government. As is evident from Wu Rui-wen’s account, post-martial law led to a period of relatively open

²⁶ Interview with author, March 3 2023.

²⁷ Ibid. Here he refers to the 1989 Tiananmen incident in Beijing, discussed below.

communication, where it was possible to receive letters from mainland China on a large scale. Avenues for cross-strait travel and phone calls also opened up, with Hong Kong as an intermediary to relay these various forms of communication.

Even these roundabout routes of exchange were significant in relation to the previous four decades of restriction. To illustrate the extremity of these prior restrictions, Michael Szonyi discusses the case of Wu Cai-sang, a young man living on Kinmen in 1949. As the home of the loudspeaker “walls of sound,” Kinmen is geographically only 6.2 miles from Xiamen in mainland China, but 116 miles from Taiwan. Prior to 1949, travel between Xiamen and Kinmen was frequent and fluid, and so on October 17, 1949 Wu Cai-sang traveled to Xiamen on a routine errand to buy cooking oil for his mother. However, while he was in Xiamen conflict intensified, resulting in the ROC claiming control over Kinmen. Szonyi writes that by the time Wu had “completed his task that afternoon, ferry service had been cut off. Stranded in Xiamen, he was unable to communicate with Kinmen even by letter or phone for decades,” even though the two locations are visible across the strait (Szonyi 2013, 74). By 1987, Wu Cai-sang was finally able to travel back to Kinmen to see his family, albeit on a roundabout route through Hong Kong and then Taiwan.

In addition to letters and visits, phone calls provided new channels for communication. These phone calls were especially valuable for people trying to connect with long-lost family members. In a famous scene at the end of the Taiwanese film *Banana Paradise* (香蕉天堂), released only two years after the end of martial law in 1989, a man and his son from Taiwan call the man’s father (the son’s grandfather) in mainland China for the first time in decades. Upon hearing his father’s voice, the man bursts into tears. The irony of the moment, however, is that in the chaos of the KMT’s move to Taiwan and the events of the White Terror period the man has taken on various fake names and identities. He is not related by blood to the son or the grandfather. While the man is fully aware of this fact, he still allows himself to be caught up in the moment, tearfully referring to the man on the other line as “dad.” Further exaggerating the irony, the son also is not related by blood to the grandfather, though he remains unaware of this fact (K.H. Chen 2010, 137; Tung 1989). Years of turmoil and restricted communication culminate in this single phone call, which at once demonstrates the opening of new communication channels and the degree to which now irreparable miscommunication and confusion have built up over time.

Despite the associated challenges, the new avenues for exchange that opened in the late 1980s were unprecedented in four decades of cross-strait relations. These new opportunities came to fruition not just in the Taiwan Strait, but also around the globe in what might be considered a worldwide crumbling of walls. 1989 marked the collapse of the Berlin Wall. The dissolution of the Soviet Union in 1991 can be understood as marking the end of the Cold War.²⁸ This was also a time when the reach and form of media were rapidly changing. Far more than simply reporting on events, the media long been conceptualized as playing a pivotal role in how

²⁸ This understanding has been exposed by some scholars to be a Western-centric viewpoint. Kuan-hsing Chen emphasizes that the “cold war structure in East Asia have weakened, but by no means dismantled” (K.H. Chen 2010, 119). Similarly, Jini Kim Watson argues that the end of the Cold War is often thought as marking the end of totalitarianism even while decolonial movements have continued through the Global South: “Today’s resurgence of ethno-nationalist autocrats is often understood as a “return” of something that has long been in abeyance” (Watson 2021, 8).

events unfold.²⁹ In the context of the Soviet Union, anthropologist Alexei Yurchak demonstrates that between 1987 and 1988, the circulation of newspapers and literary journals jumped as much as tenfold, with the most popular publications frequently selling out at newsstands. Yurchak argues that this rapid increase in the spread of information created an incongruity wherein many people in the Soviet Union harbored the belief the socialist system would last forever, but at the same time were not surprised when it ended, in part due to changes in the media environment that anticipated this event (2005, 2–3).

In the Taiwan Strait, the opening of cross-strait communication channels resulted in what has been called a “mainland media craze” (大陸新聞報導的熱潮) in Taiwan. Historian Timothy Weston suggests that this craze can be traced to September 1987, a few months after the end of martial law, when two Taiwanese reporters defied restrictions and traveled in China for two weeks (first obtaining visas by way of Japan). This visit became “a major news event in and of itself,” signaling that “the door for Taiwanese journalists to visit China had thus been opened” (Weston 2013, 215). Newspapers then rushed to meet the demand for information about mainland China. Over two years, from 1987–1989, the total number of registered newspapers jumped from 31 to 125. With the legalization of opposition parties to the KMT government, the media was no longer limited only to official state narratives but could instead represent a variety of political views (Weston 2013, 215). The diversification of political views represented in newspapers paralleled trends in radio broadcasts. While the KMT government still dominated broadcasting resources and was slow to dole out broadcasting licenses, dozens of underground radio stations emerged and competed over the airwaves, aiming provide a platform for diverse political views and express the interests of underrepresented groups, and taking advantage of a novel environment where dissent was no longer illegal (Ke 2000, 417).

At the same time, even government-backed radio was changing, enabling a diversity of voices to be represented on the airwaves. In the context of Voice of Asia, mainland Chinese listeners’ perceptions of Taiwan were no longer limited to broadcasters’ addresses, but were also informed by direct exchange through letters, phone calls, and in-person visits. Broadcasters’ perceptions of their mainland Chinese listeners were no longer limited to speculation based on occasional correspondence but were instead informed by frequent interaction and exchange—by letters so abundant it was impossible to finish reading them all. At a time when shifting political alliances and an evolving media landscape led to changes around the world, multimedia exchanges surrounding radio broadcasts contributed to transformations in what can be called the cross-strait socio-political imaginary.

The idea of the social imaginary is useful here because it encompasses not individuals’ perceptions, but instead collective understandings of how individuals fit within broader social and political positionings (James 2019, 40; Taylor 2004, 156). While radio has been discussed as playing a central role in the formation of the “imagined communities” that constituted the basis of nation-states, here a socio-political imaginary provides a way to consider how people conceptualize themselves in relation to others without the limitation of vexed national boundaries

²⁹ This idea references central tenets of media studies, such as Kittler’s famous assertion that “media determine our situation,” as well as McLuhan’s that “the medium is the message,” suggesting that the ways in which information is recorded, transmitted, and processed influences human perception of societal structures (Kittler 1999, xv; McLuhan 1994). See also (McCombs and Shaw 1972).

(Bessire and Fisher 2012, 23; Ginsburg 2012, 272; Anderson 2006). Social theorist Paul James explains, “Ideas are beliefs expressed by individuals; ideologies collate ideas as ‘comprehensive belief systems composed of patterned ideas and claims to truth;’ and imaginaries are convocations of the social whole that frame different ideological contestations” (James 2019, 42; Steger 2009, 5). The social imaginary is then a space to grapple with how belief systems are positioned in relation to other belief systems. Social theorists Craig Browne and Paula Diehl expand on the social imaginary to specifically suggest a political imaginary, or a “collective structure that organizes the imagination and the symbolism of the political, and therefore organizes the instituting process of the political” (Browne and Diehl 2019, 394). They suggest the political imaginary as an analytical tool for social science research on the political. In this dissertation I rely on a broad definition of the political, with politics not limited to official state actors but also including the ways that laws, regulations, and policies play out in the daily lives of individuals. This definition then blurs the line between the social and political imaginary.

The framing of the socio-political imaginary in the context of the Taiwan Strait provides a means to examine how ideologies shifted in relation to each other during the pivotal period when communication opened in the late 1980s. Cross-strait imaginaries during this time were informed not only by the media, but also by growth in business relations. In Taiwan, the 1980s were a time of rapid economic growth, known as the “Taiwan Miracle,” that was “jointly engineered” by “nationalist authorities and US-trained technocrats” (Yeh 2013, 6; Gold 1986; Greene 2008). Taiwanese businesspeople (台商) could increasingly travel to and set up shop in mainland China, taking advantage of low-cost production in a developing economy. Political scientist Shelley Rigger writes, “From 1987 to 1995, Beijing-Taipei relations saw unprecedented amity. After decades of civil war, the two sides were negotiating. For the first time in a century, Taiwanese were traveling to the mainland regularly, setting up businesses and investing huge sums” (Rigger 2011, 146; 2021).

That Taiwan Semiconductor Manufacturing Corporation (TSMC)—the world’s most advanced manufacturer of computer chips—was founded in 1987 also indicates the central role that Taiwan would play in developing and manufacturing computer hardware, foreshadowing a shift away from radio as a primary means of cross-strait communication as new possibilities emerged with the adoption of the internet. The founding of TSMC was in many ways a continuation of Taiwan’s centrality in the manufacturing of transistor electronics. In the 1960s, Taiwan’s electronics industry experienced rapid growth, with Taiwan becoming a central manufacturer for transistor electronics including radios (Jones 2020, 151). The founding of TSMC symbolizes the growth and expansion of that industry, with the company going on to produce the world’s smallest transistors to power advanced computer hardware and mobile devices. TSMC’s microchips would accelerate the spread of computer technologies worldwide, just as Taiwan’s transistor electronics production enabled the spread of transistor radios in previous decades.

1987 was then merely the beginning, as the widespread adoption of communication technologies continuously opened new possibilities for communication. These shifting conditions across various sectors opened the door to a period of time in which people’s imaginaries of their neighbors across the Strait were now informed not only by cultural history, generational memory, education, and state-sponsored ideology, but also by business relations, on-the-ground reporting, and direct exchanges through letter, phone, and visits.

Multimedia, Multidirectional Broadcast Radio

Broadcast radio is often considered to be a one-directional medium: the announcer sends a signal out into the world where it is received by listeners who are not required to respond. This is not to say that listeners are passive receivers: in fact, “listening publics” have been conceptualized as playing active roles, with radio long serving as a central space through which listeners can develop solidarity, resistance, and empowerment among oppressed groups.³⁰ While listening publics constitute an essential part of broadcast radio, the definition of broadcast radio most often is not predicated on interaction between broadcasters and listeners. In the context of Taiwan’s broadcast radio prior to the 1980s, there were only limited exchanges between broadcasters and listeners, with broadcast radio considered a tool for the spread of propaganda and psychological warfare.

However, during RTI’s golden era Taiwan’s central broadcasting changed in tone. Having already adopted a more cultural approach to programming, broadcasts now centered on interactions between broadcasters and listeners. Broadcast radio encompassed not only transmissions to portable radios across the Taiwan Strait, but also the letters, phone calls, and in-person visits through which listeners responded to these broadcasts. Conceptualized in this way, broadcast radio can be defined as a multimedia assemblage with multidirectional avenues for communication. Drawing on network theory, broadcast radio has traditionally been considered a star network, where a central point sends information out to multiple points. Here, the star network has bidirectional edges: information is not only sent out to multiple points but is also sent back from multiple points to the center (Figure 4) (P. Singh and Verma 2015).

³⁰ I draw on media theorist Kate Lacey’s conceptions of “listening publics” (Lacey 2013). Frantz Fanon discussed radio as a tool of resistance as early as 1959 in relation to the decolonizing role of radio in the Algerian war, where radio “inaugurates a revolutionary mode of active listening” (Allan 2019, 189; Fanon and Chevalier (trans) 1965; Baucom 2001).

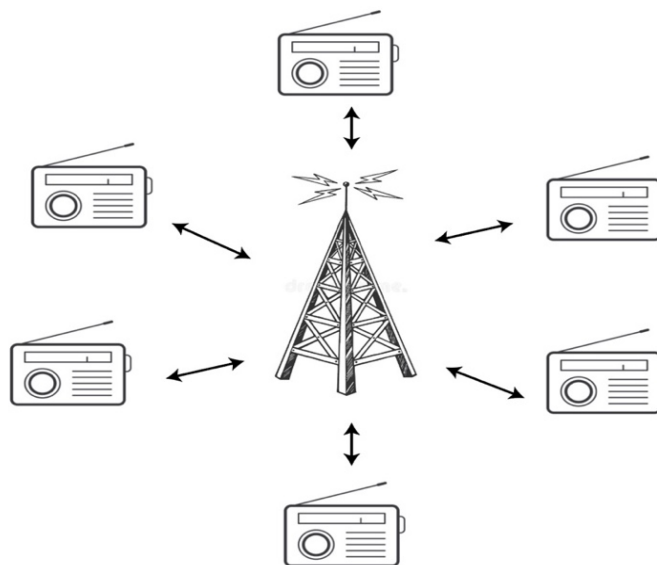


Figure 4. A star network with bidirectional edges. The two-sided arrows indicate that information is not only sent out to multiple points but also is sent back from multiple points to the center. Figure courtesy of Satcher Hsieh.

Conceptualized in this way, radio as a site for exchange begins to align with the aspirations of certain early radio scholars. Communications professor John Hochheimer notes that although radio was initially structured to serve political, military, and economic interests, figures like Walter Benjamin, Bertolt Brecht, and Theodor Adorno sought greater involvement of listeners. He writes:

Walter Benjamin spent several years attempting to develop radio forms in which listeners exercised greater control over form and content...Bertolt Brecht, too, wrote essays...in the 1920s and 1930s...in which he tried to develop ‘art for radio and radio for art,’ bringing listeners into more direct responsibility for production and program choice...Theodor Adorno constructed an extensive memorandum speculating about the...ways in which listeners could become more actively engaged with the medium (Hochheimer 1993, 473–74).

Anthropologist Lucas Bessire has suggested that two-way radio, unlike traditional broadcast radio, fulfills the visions of these early intellectuals by offering a truly interactive communication form. Two-way radio offers a “walkie-talkie” dynamic where back-and-forth responses are expected, creating a conversational exchange between two parties that are at once producer and consumer. In the context of RTI’s golden era, the emphasis on multidirectional exchange through broadcast radio then blurs the distinction between broadcast and two-way radio, where listeners become producers, broadcasters become consumers, and interactions are characterized by conversational multiplicity (Bessire 2012, 213).

There are, of course, limitations here. Bessire's suggestion that two-way radio is a "horizontal participation platform" does not fully transfer to the context of RTI's golden era of broadcast radio (2012, 200). The relationships formed through Voice of Asia encompassed age, power, and fame differentials between broadcasters and listeners, perhaps falling on a diagonal axis rather than a horizontal one. This power differential was less the case in the context of Taiwan's boom in the underground radio scene at the time, where call-ins from listeners could take up the vast majority of programming time on these stations.³¹ Even while community radio sought to provide an alternative to state-run radio, the participatory nature of Taiwan's state-run radio in fact echoed that of community radio. Despite opposing political goals, the nature and structure of radio interactions began to sound similar. Broadcasters and listeners collaborated and engaged in meaningful multimedia and multidirectional exchanges, and in the process collaborated in the co-creation of a shifting cross-strait socio-political imaginary. These interactions led to distinct relationships formed through and around sound, and in turn a new model for what cross-strait relations could sound like. Whether state-owned or underground, media was becoming decentralized. Radio was no longer exclusively a platform for a broadcaster's voice, and instead could include a variety of voices.

It was therefore not only the popular music and culture-focused broadcasts that define the golden era of RTI, but also the increased opportunities for exchange through letters, phone calls, and visits. While the popular music content of the broadcasts was a big attraction for the mostly teenaged listeners in mainland China, recent accounts of this period from former broadcasters and listeners now largely emphasize the interactions and exchanges. Even more important than the content itself were the exchanges surrounding this content and the relationships formed. Indeed, in describing the typical letters he would receive, broadcaster Wu Rui-wen explained that the content of most of the letters consisted of trivial details about listeners' lives. However, what the letters had in common was the expression of the hope that they would receive a response:

For the content of these letters, of course I've kept the good ones, but actually they are all quite simple. Everyone is most concerned about themselves, so they start from themselves. First, they express that they were so happy to hear your program. Then that they long for the rich world of Taiwan. Then they share how their lives are, their academic situations, their emotional world, their social circles. Then that what they look forward to the most is the weekly 'Listener's Mailbox' segment, where we can respond to their letters on air, and let them hear the letters. When their letters receive a response, it gives them a sense of being valued.³²

[這些信的內容，當然好的信件我基本上都會保留，其實都很簡單，每個人最關心的是自己，所以他都是從自己出發。第一個他表達他聽到你們的節目覺得很高興。那麼對於臺灣的一些豐富的世界他們很嚮往。然後他們的生活狀況是如何？他們的學業狀況是如何啊？他們的感情世界，他們的交友世界。然後他們最期待的就是每個禮拜的這個聽友信箱，我們在空中能夠恢復他們的信件，把他們聽到信件。這個在信件被恢復就有一種受重視的。]

³¹ Ke writes that on one 30-minute program, 25 minutes consisted of listeners calling in (Ke 2000, 418). The growth of community radio in Taiwan around this time could be conceptualized as part of a global trend where community radio was thriving around the globe (Fisher 2012; Scifo 2014).

³² Interview with author, March 3 2023

From the listeners' perspective, Cindy (who listened to Voice of Asia from mainland China as a teenager) said that one of her main motivations for writing letters was the feeling she would get when they were read out loud on air:

There was a special program where they would respond to listeners' letters, and if they responded to your letter, you would feel good about it, so I would write them letters. The content of the letters was just responding to a program I heard, or something happening in my life, that kind of thing.³³

【 呃，就會聊一些聽節目的，因為其實當時其實也沒有想，是因為他們其實會有一個節目頻道就是回聽，有的信覺得說，哎，被被主持人念出來自己的性感覺喔，好好好好有一種好像談不錯的感覺，所以自己也會寫。他寫的內容就是哎，聽廣播的感受啊。或者是，呃，寫一些就是在自己周遭發生的一些事情很有趣的事情跟他們做一些分享，就這樣對。】

Described here is the process through which written letters were remediated to voice as broadcasters read listeners' letters out loud on air. While interactions surrounding broadcast radio were multimedia, this process indicates the continued centrality of sound and voice in this context. Once read aloud, these exchanges took on heightened meaning for listeners as they received responses and validation from broadcasters for their written communication. By vocalizing listeners' letters, broadcasters demonstrated their deep engagement with their audience, literally embodying the listeners' voices through these exchanges.

Indeed, while Cindy could have listened to cassette tapes to access popular music or chosen mainland Chinese radio broadcasts for entertainment, she preferred Taiwanese broadcasts for the sense of personal connections:

At that time, in terms of media, you had to go buy cassettes. And then for mainland radio I didn't listen as much, because Taiwan's radio gave me a sense of being close to others. So it was easier to listen to overseas broadcasts.³⁴

【 那個時候我覺得資訊的部分來講，你要去買卡帶，然後如果大陸的話，其實廣播的內容，我聽得比較少，因為臺灣的其實那個時候給我感覺蠻接近彼此的那種感覺所以比較會容易去聽海外的廣播。】

These multimedia exchanges took on meaning less through their content and more through the very acts of writing, listening, and responding. As media historian John Durham Peters writes, these acts of communication were “existential, not informational. The two parties had nothing to say, but everything to mean” (Peters 2015, 14).³⁵ Through these acts of communicating, listeners and broadcasters continuously asserted their presence to one another. While the content of the exchanges may have consisted of trivial discussions of daily life, the practice of cross-strait

³³ Interview with author, March 10 2023

³⁴ Interview with author, March 10 2023

³⁵ Peters' discussion references Amos Oz's book *A Tale of Love and Darkness*, in which Jewish relatives in Jerusalem and Tel Aviv in the 1930s and 1940s would write letters to arrange phone calls. The phone calls would consist of trivial exchanges, and they would then hang up and continue they cycle of writing letters to arrange the next phone call. The need for the phone calls was not to exchange information, but rather to “hear each other's voices” and express their aliveness at a time when the fate for Jews in Europe “hung by a thread.” The call was then existential, not informational (Peters 2015, 13–14).

exchange during this period was novel, personal, and inherently politicized. Conceptions of the cross-strait socio-political imaginary were then not only shaped by the content of broadcasts, but also by the sounds of each other's voices and the mutual feelings of engagement and connection facilitated by the act of communication.³⁶

Writing letters was one aspect of the multimedia exchanges through which broadcasters and listeners developed interpersonal connections. These letters also served as a measure of success of the program, where RTI could monitor listener engagement based on how many letters they received each month. Even more, the number of letters received provided a measuring tool for the climate of cross-strait relations. In 1987 with the end of martial law in Taiwan, Wu Rui-wen said that the letters came in droves. In 1989 with the Tiananmen incident, the letters abruptly stopped, before starting up again a few months later.³⁷

The 1989 Tiananmen incident in mainland China was the culmination of growing activist movements throughout the 1980s. Rigger explains, "The 1980s were a period of unprecedented political opening in the PRC. Where Mao's China had been a nation of surveillance, propaganda, and control, under Deng Xiaoping, Chinese were permitted to speak and think for themselves to an extent not seen since before the PRC's founding in 1949" (Rigger 2021, 38). While Deng Xiaoping's economic reforms catalyzed significant growth, they also led to undesirable effects such as inflation, corruption, and a widening income gap, fueling public discontent particularly among students and workers. Growing activist movements culminated in the Spring of 1989 with the Tiananmen incident. Originally a response to the death of reformist leader Hu Yaobang, the memorial in Tiananmen Square in Beijing transformed into a broader movement demanding political reforms and greater freedoms. These protests gained global media attention and lasted for weeks until the government ordered military intervention, violently ending the protests and leaving an unknown number of people dead (Rigger 2021, 39).

This event, while condemned internationally and by the KMT government in Taiwan, in fact in many ways served to strengthen cross-strait solidarity and business relations. The event's live broadcast on RTI played an important role in how it was interpreted in Taiwan. Sociologist Ming-sho Ho writes:

On the evening of 3 June, a mass rally called 'Connection of Blood Veins and Singing Across the Strait' (血脈相連兩岸對歌) was held in the Chiang Kai-shek Memorial Hall. The highlight was a telephone call connecting music performers in Taipei and Beijing. Around midnight, the news that People's Liberation Army soldiers had opened fire on demonstrators was instantly broadcast over the telephone, and the communication was abruptly terminated. The event quickly turned into a mass mourning, lasting until the next morning (Ho 2022, 94).

This account demonstrates that Taiwanese people experienced the Tiananmen incident in real time through phone call and live radio broadcast—something unthinkable just three years prior.

³⁶ Here I am alluding to discussions of "phatic" behavior, or communication for the sake of connection rather the content of exchanges. This idea can be traced to Bronislaw Malinowski and has been taken up more recently to discuss social media interactions which facilitate the act of connection over in-depth exchange (Gradinaru 2018; Malinowski 1923; V. Wang and Tucker 2016).

³⁷ Interview with author, March 3 2023.

The Tiananmen incident evoked sympathy and solidarity in Taiwan, where both Taiwanese activists and the KMT government mobilized to express support for the mainland Chinese activists, albeit with differing motivations. The event and its associated international attention were bad press for the CCP, and the KMT government capitalized on this international attention to legitimize its own political aims. From this event, the KMT “contended that mainland compatriots were no longer supportive of the communist regime,” realizing that “international condemnation of Beijing appeared to lend credibility to the Kuomintang’s outdated claim as the sole legitimate government of China” (Ho 2022, 94). After June 3, the KMT government continued to host large-scale rallies where participants sang patriotic songs, reinforcing “a message of national unity” between mainland China and Taiwan (Ho 2022, 100). Taiwanese businessmen looking for economic opportunities also took advantage of the situation. After the incident, international governments imposed economic sanctions, and the Chinese government pulled back from its reform and opening policies. However, Rigger notes that “the vacuum that formed when other foreign investors fled might have broken China’s export drive but for the Taiwanese.” Taiwanese businessmen rushed to fill the gaps, taking advantage of this brief interruption of China’s economic rise to expand their investment in mainland China (Rigger 2021, 40–41).

On the other hand, Ho writes that student activists and dissident intellectuals in Taiwan criticized the KMT’s handling of the Chinese democratic movement, maintaining that solidarity with students in Beijing was not predicated on Taiwan and mainland China belonging to one nation. They also suggested that this incident could be compared to the February 28, 1947 (228) incident in Taiwan, when the newly arrived KMT authorities violently cracked down on anti-KMT resistance. While these perspectives remained a “minority voice” at the time, they indicated a budding activist movement in Taiwan. In fact, the Tiananmen incident can be thought to have inspired a parallel student protest movement in Taiwan in March 1990, the Wild Lily Movement (野百合學運), with protestors aiming to expedite political reforms. Ho points out, “Students were emboldened to take unconventional action because they knew that the government could not afford the cost of applying the Tiananmen solution which it vehemently condemned” (Ho 2022, 95). Even as this sense of solidarity permeated the cross-strait imaginary, attitudes in Taiwan were shifting, with people beginning to identify less with a Chinese identity and more with a unique Taiwanese identity. While the late 1980s harbored opportunities for cross-strait collaboration and solidarity, a decade later these visions became farther out of reach.

The Tiananmen incident then strengthened both cross-strait business relations and activist solidarity, while only leading to a brief pause in the newly opened opportunities for communication and exchange. Once the dust had settled, the letters came flowing back to RTI. Soon after, the Voice of Asia broadcasters joined Taiwanese business travelers in crossing the strait and began in-person visits to mainland China for what they later called “listener meetups” (聽友會). The broadcasters were colloquially known to listeners as the “Four Heavenly Kings (四大天王, see Figure 5), and Wu Rui-wen said they were greeted by fans in mainland China as if they were superstars:

Of course, you would feel like a star, you would feel a sense of vanity. Let me tell you, when we went to Fuzhou, we had to take a flight and go through customs. The customs officers saw our Mainland Travel Permits, our passports, and shouted, ‘Ah, who is this?’ It was like they were calling out the names of Taiwanese stars. Even the customs

officers recognized us. You would feel a sense of vanity, I admit that. But putting that aside, what I felt most deeply was that they heard the messages we delivered. And I believe the most important thing was that the messages we brought to them opened a very bright window in their lives, giving them a greater understanding of the outside world.³⁸

當然你會有一種你是一個明星的感覺，你會有虛榮感。我跟你講，我們當時到福州去啊，我們要坐飛機進去海關。那個海關人員吶，看到我們的這個臺胞證，我們的護照，就高喊，‘啊，這個是誰？’ 比方這好像在喊臺灣的明星一樣。連海關都認識我們。這些你會有虛榮感，這個我坦承講。不過這個都放一邊了，我覺得最最受深刻的就是我們傳遞的訊息他們有聽到。然後呢，我覺得最最重要的是我們帶給他們的訊息，使得他們的這個生活呢打開了一扇很光明的窗，他們對於外面的世界呢，有了更多的認識。



Figure 5. The “Four Kings” (四大天王): Voice of Asia broadcasters Shen Wan (沈琬)、Xie De-sha(謝德莎), Lin Xian-zheng(林賢正), and Wu Rui-wen (吳瑞文). Photo courtesy of the Facebook group 曾經的美好-亞洲之聲 (<https://www.facebook.com/photo.php?fbid=601290811472504&set=pb.100047747722092.-2207520000&type=3>).

As a listener, Cindy said she put in considerable effort to attend these meetups. In the earlier years, Cindy once attempted to go to a meet-up accompanied by her mother. They took the train two hours to get there, but they got the time and place wrong and never found the event. 10 years later, she went to a listener meetup in Shanghai with a friend she had met through their mutual interest in Voice of Asia. Cindy said they stayed there for two or three nights. At one of

³⁸ The four kings are Shen Wan (沈琬)、Xie De-sha(謝德莎), Lin Xian-zheng(林賢正), and Wu Rui-wen (吳瑞文). Interview with author, March 3 2023

these meetups, Cindy brought a gift for Wu Rui-wen related to one of her favorite programs at the time:

The program that impacted me most is this kind of program that doesn't exist anymore. The host would read out the lyrics one line at a time, and we would carefully write down the lyrics. All those songs were new songs that hadn't been released in mainland China at the time. Almost everyone learned to sing the new songs because of this program. The songs I can sing even now are still these old songs. The program was called "You Sing With Me." The host would read the lyrics line by line, and the listeners in front of their radios would write them down. We just assembled one notebook after another like this. I remember I used two notebooks, and when we had a listener gathering, I gave one of them to my friend Wu Rui-wen.³⁹

[我印象最深的就是這種節目，現在都不會產生，就是主持人拿著歌詞念，念一句念很久，然後我們真的就是很認真地把歌詞抄下來。那些歌全部都是新歌，然後那個時候還沒有在大陸發行的歌，所有的人幾乎就是因為聽了這個所有的新歌都會唱。包括我現在會唱的，其實都是這些老歌。這種節目的名字就叫你唱我和。就是一句一句的念那個歌詞，讓收音機前的那個聽眾把那個記下來。我們好多人呢，都集結了一本一本。然後我記得我也有用了兩本，那個時候開聽友會，我就拿給吳大哥了。]

The dedication of listeners was essential to the organization and success of the meetups. This was especially the case during the earlier visits, when the events were not designated as official meetings to avoid drawing the attention of the authorities. Wu Rui-wen explained:

Because they didn't have freedom, it would be problematic if there were gatherings of around 50 or 100 people. So we cautiously went in. When we went in, we didn't say we were going to hold a listener meeting. We would just say on the program, 'On a certain day of a certain month in a certain year, so-and-so and so-and-so will be at a certain hotel in Fujian, and we are going there to travel and meet old friends.' We would usually say something like that. So by saying that, even before we arrived many people would call the hotel, and then of course, they would wait outside the hotel wanting to meet us. At the time, we would see how many audience members actually came, and we might find a place in the hotel to meet and chat with them. Among them, there were some who had often written to us in the past, so they could help us answer the phones at the hotel. So, our format was generally not to explicitly say we were holding a listener meeting. But the first time we held a listener meeting in mainland China, the [listener meet-up] sign wasn't made by us but by the listeners themselves. The listeners made it on their own initiative and managed everything related to it.

[因為他們是沒有自由的，大概可能 50 人 100 人以上的這個機會的話呢，是會有問題的，他們會被調查。那我們是試探性地進去。我們進去的時候呢，基本上我們沒有講我們去開聽友會。我們一定是在節目裡面講說啊，'某年某月的某一天，某某人跟某某人我們會到福建的那一個飯店喔，我們要去旅遊，見見老朋友，'大概我們都是這麼講。那么這麼講我們其實人還沒到，很多人就打電話到那個飯店裡面去，然後他們當然會到飯店外面留守啊，想跟你見上一面。那個時候我們再看實際參與到來的這些受眾有多少，我們也許在飯店找一個地方，就跟他們就地大家見面就聊聊。那么這當中當然也有一些過去的時間裡經常會給你寫信的，所以這些人就可以幫忙你在飯店裡面接電話。所以我們的形式大概就是不會去講開聽友會。可是第一次我們中國大陸去

³⁹ Interview with author, March 10 2023.

開聽友會的時候，[聽友會的]牌子不是我們自己做的，是聽眾自己幫我們做的啊，就是聽眾自己做的，自動自發去做相關的處理。]

As his account indicates, listeners were heavily involved in these visits, playing a central role in organizing and coordinating the events, and even branding the visits as “listener meetups.”

Later, the broadcasters also called on listeners to help coordinate collective letter delivery. This was because sending mail locally in China was less expensive than postage from mainland China to Taiwan. A listener volunteer would set up a local mailbox and inform the Voice of Asia broadcasters of the address. Then the broadcasters would announce the address on the radio. When they arrived in mainland China for a listener meet-up, the listener volunteer would deliver all of the letters.⁴⁰ This arrangement indicates that the relationship between broadcasters and listeners was multidirectional and mutually beneficial. While the listeners depended on the broadcasters to share information, inform them about new music, and establish the foundation for a listening community, the broadcasters depended on listeners to contribute their thoughts and feedback and to coordinate events.

The Beginning of the End of the Golden Era

The dedication and participation of listeners is particularly noteworthy considering the varying degrees of political risk involved. While cross-strait exchanges during the late 1980s through 1990s were safer than in previous decades, they were still often under scrutiny of the mainland Chinese authorities. Wu Rui-wen emphasized that broadcasters were cautious not to endanger their mainland Chinese listeners, deliberately avoiding political discussions and hoping that listeners would refrain from addressing political topics in their letters and exchanges. However, even the act of establishing relationships between Taiwanese broadcasters and mainland Chinese listeners had political implications, since these broadcasts provided listeners with a means to move beyond mainland China’s closed-off media environment.

To illustrate the risks involved, Wu Rui-wen remembered that one of the listeners helping Voice of Asia to collect letters faced some consequences:

There was one time we had a listener in Nanjing set up a mailbox for us. On that trip, we traveled to various provinces in mainland China and held listener meetings, moving south. Just before we arrived in Nanjing, this listener suddenly disappeared the day before our arrival. He had previously reported to us that he had collected about two to three thousand letters. But after we arrived in Nanjing, this friend was missing. We couldn’t find him. Other listeners told us that he might have been temporarily detained because of collecting letters for us. Later, listeners took us to his home based on the address we had. We wanted to find him and give him our regards, but in fact, they already locked the door, so we couldn’t see this listener. So this was a big regret for me.

⁴⁰ To further illustrate the extent to which cross-strait communication was for the most part safe and open during this time, Wu Rui-wen discussed an exchange they did with a mainland Chinese station, Voice of the Strait (海峽之聲). Voice of the Strait contacted Voice of Asia after realizing they had a large audience to ask the Voice of Asia broadcasters to be guests on their program. Afterwards, Voice of the Strait also received and forwarded letters to Taiwan for Voice of Asia (Interview with author, March 3 2023; L. Li and Gregory 2009).

But I want to say that during this process, I went to the mainland over ten times, and if we distribute over 10 years, it was generally quite smooth. However, as mainland China became stronger and more prosperous, its control also became stronger, so our space to operate became smaller and smaller.⁴¹

[有一次我們在南京有一位聽眾幫我們設了信箱。那一路呢，我們到大陸啊，去這個省份，那個省份開完聽友會一直往南走的時候呢，居然在前一天，我們要到達南京的前一天，這位聽眾消失了。他之前跟我們匯報，他大概收了可能有兩三千封的信件左右了吧，可是到了南京之後呢，這位朋友不見了，我們找不到這個人了，那聽眾告訴我們說這個人可能因為幫我們收信的關係呢，暫時被隔離了。後來聽眾還帶我們，我們按照那個地址去這個聽眾的家，我們要去找他，要對他致意。但是呢，事實上，他們已經門戶深鎖了。我們也遇不到這個聽眾。所以這是我蠻大的一個遺憾。但是我要講的就是說，在這個過程裡面，其實我去了這個十多趟，如果說用10年來劃分的話，其實都蠻順暢的。可是越到後來中國大陸越強大越富強了，它的那個控制力就越來越強了。所以呢，我們要發揮的空間就越來越小了。]

Here Wu Rui-wen notes a shift in the political climate in the mid-1990s, after which listening to Taiwanese radio in mainland China became riskier due to increased cross-strait tensions. Cindy also remembers various stages of interaction with the mainland Chinese authorities. When she was young in the late 1980s, many of her letters were confiscated before reaching the hands of broadcasters. Skipping to a decade later, restrictions against listening to Taiwanese radio were enforced more thoroughly, and radio signals were more likely to face static from jamming:

At that time [in the late 1980s], although it was open relatively speaking, there were still restrictions. For example, when I sent letters to Taiwan, actually many of the letters written when I was young were never received. Probably none of them were received. Later, when I was working, I sent letters again, and they were finally received, and they would read them out. Even during middle school when I wrote, they received a few letters. And then the mainland's relevant authorities would have a chat with you. That was just the lightest, lightest reminder, telling you not to do certain things. Because they would actually review the contents of the letters you sent. Then later after a bit of time when I had started working, the supervisor had a talk with me and my dad said to put away the radio. So during that period, I put it away. There was no way to listen. Sometimes I would secretly listen. But actually, in terms of the shortwave radio reception, it was not very good, it had static.⁴²

[那個時候其實雖然是開放，但是相對來說其實還是會有一些封閉的狀態，包括因為我寄台灣的信件，那其實我那個時候，小時候其實很多信寫的就收不到，可能他們都收不到，然後後來工作的時候也有寄，那終於有收到，那他們會唸出來。大概國中的時候吧，我就寫的時候他們都有收到幾封。然後主管機關就會找你聊一下。那個是最最最最輕的那個是很輕很輕的提醒你說，不要去做這件事情，不要怎麼樣。因為你寄信的內容他其實都會去看。那後來有一段時間已經參加工作了之後，那個主管找我談話，然後我爸就說那把收音機就收起來吧，所以我們家後來那段時間，其實我就收起來了，沒有辦法，就有時候偷偷聽一下。但其實因為短波的收音的其實效果來講，其實都不會很好，他是會有雜音的。]

⁴¹ Interview with author, March 3 2023

⁴² Interview with author, March 10 2023

After being reprimanded at work as a young adult at the time, Cindy felt she hadn't done anything wrong. She was simply listening to the radio broadcasts that she loved so much. Looking back on it now, however, she said that she understands her actions in the broader context of shifting cross-strait relations.

Indeed, by the mid-1990s the brief period of relatively open cross-strait interactions had begun to wane. During this time, the Taiwanese democracy and independence movements grew, culminating in Taiwan's first direct presidential election of Lee Teng-hui in 1996.⁴³ Lee's election signified a move toward a unique Taiwanese identity, with his platform emphasizing that "everyone living on Taiwan, regardless of when they or their ancestors had arrived, was part of a single 'community of shared fate'" (Rigger 2011, 147). Lee adopted an international strategy of "pragmatic diplomacy," which moved away from the "Chiang-era insistence on formal recognition" toward building unofficial international relationships. Lee's presidential campaign culminated in a controversial visit to Cornell University in 1995. This visit was perceived by mainland China as a betrayal, and in response Beijing intensified military pressures. In what is known as the Third Taiwan Strait Crisis, the People's Liberation Army conducted extensive military drills including missile launches and naval exercises near Taiwan, significantly escalating tensions in the Taiwan Strait ahead of the election (Rigger 2011, 145–46).

Though the 1995-1996 Taiwan Strait Crisis represented the renewed volatility of cross-strait relations, it did not for the most part disrupt the growing opportunities for cross-strait travel, business, and communication. It is however representative of a qualitative shift. Rigger writes, "Beijing's military intimidation arrived in Taiwan's freewheeling atmosphere like a SWAT team at a fraternity party. It showed Taiwanese that the newfound friendship between the two sides was shallow and fragile, and it forced them to face a painful reality: the PRC was serious about unification." For Taiwan's growing independence movement, this meant that they would "have to defeat not only the KMT but also the Chinese Communist Party—and the People's Liberation Army" (2011, 146).

As mainland China became "stronger and more prosperous" (越強大越富強) (in the words of Wu Rui-wen), the mainland Chinese government increasingly had the resources to jam radio signals from Taiwan, resulting in the static and noise that Cindy heard on the airwaves. Whereas tuning in to enemy radio had for decades provided a way to circumvent restrictions on communication, now the regulatory mechanisms were finally catching up to the communication technologies. At the same time, the internet was emerging as a new frontier for communication, increasingly stealing audiences away from old-school radio broadcasts, and continuing the cycle of regulatory latency in line with the development of new and as yet unregulated communication technologies.

The shift in cross-strait relations marked by the Third Taiwan Strait Crisis signaled the beginning of the end of RTI's golden era. Despite these challenges, cross-strait exchanges surrounding radio continued to thrive for a few more years even as Taiwan's state broadcasts shifted in structure. In line with increased governmental decentralization in Taiwan, in January 1998 Voice of Asia and Voice of Free China officially merged with the Central Broadcasting

⁴³ Rigger writes, "The early 1990s saw a cascade of democratic "firsts": the first election of all-new National Assembly members in 1991, the first election of an all-new Legislative Yuan in 1992, the reintroduction of directly elected Kaohsiung City and Taipei City mayors in 1994, the first direct election of the provincial governor in 1994, and, in 1996, the first popular presidential election" (2011, 79).

Station, or the KMT-founded station long housed under the Ministry of National Defense and responsible for propaganda efforts directed toward mainland China. Through this merge, CBS became an independent corporation and Voice of Free China was renamed Radio Taipei International, and then later Radio Taiwan International (Rawnsley 2000, 124–30; L. Li and Gregory 2009). The structure of Taiwan’s radio broadcasting became increasingly decentralized and depoliticized, at least nominally marking the official end of Taiwan’s psychological warfare efforts.

With this restructuring, cultural broadcasts from Taiwan toward mainland Chinese listeners pressed on. The day before the reorganization officially went into effect on New Year’s Eve 1997, Voice of Asia transmitted a special holiday broadcast live from Sun Moon Lake, a scenic tourist spot in central Taiwan. This broadcast is representative the typical sounds of Voice of Asia at the time as well as the program’s popularity, with enough sway to invite top celebrity guests and a non-stop line of mainland Chinese callers eager to spend their New Year’s Eve listening to Taiwanese radio broadcasts.⁴⁴

Listening to a recording of this broadcast in RTI’s office, I am transported back to New Year’s Eve 1997. Live from Sun Moon Lake, the Four Kings encourage listeners to call in. A high-pitched ringing ensues. The announcers answer the phone. It’s a lady from Shandong province in eastern China. They ask her to pronounce Sun Moon Lake in her hometown dialect,⁴⁵ and then encourage her to sing a song. Her shaky voice rings out across the air. The phone rings again. This time it is a friend from Sichuan. The caller introduces himself, and host Wu Rui-wen greets him, recognizing his name and voice as if he is an old friend. The caller uses the Sichuan dialect to say “Sun Moon Lake,” and then sings a song. This pattern continues, with callers from Guangdong, Fujian, Hebei, and Fuzhou. Each caller pronounces “Sun Moon Lake” differently, in their hometown dialect. RTI is not limited to a single voice or a single national language: there is a diversity of voices represented on air.⁴⁶

On the twelfth call, the caller says, “Hello? Hello?” and Wu Rui-wen immediately replies, “It’s Xu Qin!” The other announcers are impressed that Wu Rui-wen can hear the voice and immediately identify the caller, though after a decade of exchanges they all seem to know Xu Qin well. Xu Qin pronounces Sun Moon Lake in the Anhui dialect, noting that it sounds more or less the same as standard Mandarin. He begins to sing the song “Thinking of You” (一想到你呀), a 1997 hit by Taiwanese singer A-Mei (張惠妹), beginning with the lyrics “As soon as I think of you, it makes me happy,” (一想到你呀·就讓我快樂). He rushes through the performance with enthusiasm, almost yelling the lyrics. After a few minutes, the announcers say “thank you” to Xu Qin, but he keeps singing. They say they think he sounds like a rock singer. The announcers start

⁴⁴ Due to RTI’s policies, the description of this broadcast was paraphrased from a listening session at the RTI archive and not transcribed word for word. Accessed February 24, 2023.

⁴⁵ While Mandarin is the official language of mainland China and widely spoken across the country, there is a wide regional diversity of languages and dialects as well. Sun Moon Lake is a famous site in Taiwan but is especially well known in mainland China due to its inclusion in school textbooks (Rowen 2023, 59).

⁴⁶ The presence of multiple voices as well and multiple dialects or languages on air can be considered audible proof of broadcast radio as multiplicitous and representative of a broader public rather than just one official governmental voice. Tom McEnaney discusses a similar phenomenon in the context of the American podcast *This American Life*, which “attempts to create a public voice and a public emotional connection by turning the microphone over to people other than its reporters” (McEnaney 2019).

talking, but Xu Qin's singing remains in the background. They reintroduce the program as a special live broadcast from Sun Moon Lake. When they finish, Xu Qin is still singing with enthusiasm.

It is now halfway through the broadcast, and the announcers introduce celebrity singer A-Mei as their special guest for this broadcast. They answer the phone for an especially nervous caller—the first to talk to his celebrity idol today. He hesitates to sing, but with the announcers' patient encouragement he begins from a catchy interlude in the middle of "Thinking of You": "La la la la la." A-mei joins in and they sing together:

La la la la la la make your wish
 Throw a coin and wait for the moon to be full
 La la la la la la make your wish
 Throw a coin and wait for the moon to be full
 [啦啦啦啦啦 許下你的心願
 丟一枚錢幣等月兒圓
 啦啦啦啦啦 許下你的心願
 丟一枚錢幣等月兒圓]

The caller briefly stops singing, but A-Mei continues. The caller joins in again, and they continue together. Somehow, with the gentle encouragement of these radio hosts, a nervous caller finds a way to share his voice on air with the public and his celebrity idol, feeling comfortable in this environment of exchange that has been mutually cultivated over many years.

The Lasting Legacies of RTI's Golden Era

While New Year's Eve 1997 can be considered to mark the beginning of the end of Voice of Asia, New Year's Eve 2001 marked the final broadcast. By this point RTI had already begun online broadcasts, indicating both a shift in technology and priorities. The final broadcast, which was later posted online by nostalgic former listeners, was hosted by two of the four kings: Wu Rui-wen and Xie De-sha.⁴⁷ In the program, listeners call in to leave heartfelt messages about the impact that Voice of Asia has had on their lives. One listener leaves this message:

In life, there are many unexpected encounters, but meeting you was a life-changing event for me. No words can express my gratitude for meeting you. Actually, at times like this, the ones who truly bow to you are all of the listeners. The joy and hope you bring to us will illuminate our whole lives. For real! In our hearts, you are irreplaceable because of your sincerity and passion. Because we have walked through the years of youth together, I will always be your friend.

[生命中有許多的不期而遇，你們的相識卻是改變我人生的一次。相逢千言萬語都無法表達我對你們的感謝。其實在這樣的時候，真正想要一鞠躬的是我們所有的聽眾朋友，你們帶給我們的每一份歡樂與希望，都會照亮我們的一生。真的。在我們的心裡，你們是無可取代的，因為你們真誠、自然、熱情。因為我們曾經一起走過的青春歲月，我就是你們永遠的朋友。]

⁴⁷ 2001.12.31 The last complete recording of Voice of Asia's *Good Times*. 2001.12.31 亞洲之聲 你我好時光 最後一期完整錄音. Posted October 31, 2021. <https://www.youtube.com/watch?v=S3GYm5AyLIE>

Another listener calls in:

Wen-ge, Sha-jie, I am a listener from Jiangxi. Today is my first time calling you.⁴⁸ I have been listening to the program for six years but have never contacted you before. This is my first time calling, so I am a bit nervous. I am very grateful for your companionship over these six years. I sincerely hope to hear Wen-ge and Sha-jie hosting a new program soon. There is so much I want to say, but I will let this song express my feelings for you.

[文哥·莎姐·你們好·我是江西的聽友·無非·今天是我第一次打電話給你們·我聽節目有6年的時間·但是一直都沒有和你們聯繫上·今天是我第一次打電話·所以心裡有點緊張啊·非常感謝你們在這6年的時間能夠陪我一起度過·在這裡也非常希望能夠早日聽到文哥和莎姐主持的新的節目·想說的話很多·但是千言萬語化成下面這首歌送給你們。]

His lone voice rings out over the airwaves, singing the chorus of the song “Red Beans” (紅豆) by Faye Wong (王菲):

Sometimes sometimes
I can believe that everything has an end
There is time to get together and leave
Nothing lasts forever
But sometimes
I would rather hold on and not let go
Wait until the view is clear
Maybe you will stay with me to watch the flowing water

[有時候 有時候
我會相信一切有盡頭
相聚離開 都有時候
沒有什麼會永垂不朽
可是我 有時候
寧願選擇留戀不放手
等到風景都看透
也許你會陪我看細水長流]

The caller’s voice fades, replaced by Faye Wong’s 1998 recording of the same song—likely a song the listeners were introduced to through Voice of Asia. The series of calls continues until the announcers reintroduce themselves, beginning to wrap up the program. Wu Rui-wen encourages listeners to keep writing letters, with his voice cracking as he begins to cry. The hosts say, “See you next year,” and a recording of Sandy Lam’s (林憶蓮) song “Hard to Say Goodbye” (捨不得說再見) closes out the episode.

While this emotional broadcast marked the official end to Voice of Asia, the relationships developed through and surrounding the broadcasts proved to be long-lasting. In this final episode

⁴⁸ He calls the announcers by familiar names, as if they are close friends or older siblings.

Wu Rui-wen encouraged listeners to continue to write, and they did so even as emails became more common than letters, and even as social media sites Facebook and Weibo became more common than emails. In our interview in March 2023, Cindy said that every morning when she logs on to social media, she makes sure to check the posts from Wu Rui-wen. He posts essays every day in which he recalls his experiences hosting Voice of Asia and his interactions with listeners. Cindy also reconnected with Wu Rui-wen in person after moving to Taiwan as a young adult. She said that listening to Taiwanese broadcasts as a kid may have influenced her decision to move to Taiwan because these broadcasts sparked her interest and curiosity about Taiwan. Cindy said that reading Wu Rui-wen's posts has become part of her daily routine, and interacting with him and other listeners online and in-person makes her feel that they are like family. She explained (using a familiar name for Wu Rui-wen), "Wu-ge, we are already very familiar with each other. We've known each other for almost 20, 30 years now, so the distance between the host and the listener has already become that of friends."⁴⁹ [吳大哥·我們已經很熟·現在其實認識已經差不多快20·30年了·所以已經好像從主持人跟這個聽友之間的距離·大概已經變成朋友]。

In one of Wu Rui-wen's posts on Facebook, he shared a series of messages he had received from a listener through Weibo:

I am truly grateful for the internet age, Wen-Ge. I am your loyal listener from Anqing, Anhui. I am very happy that we could exchange through the airwaves on [the program] "Good Times, Me and You." Wen-Ge, back then whenever I had free time I would always listen to your two-hour program, even during meals. I never missed the Saturday and Sunday broadcasts!

There, for the first time, I heard Wong Ka Kui and Jonathan Li. For the first time, I began to understand Luo Ta-yu's music. Voice of Asia provided me with a window to understand more about the wonderful world outside!

The most legendary and timeless stories in the Chinese music world were first learned through Voice of Asia. For the first time, I heard Luo Ta-yu, for the first time I deeply understood Dave Wong, for the first time I knew Wong Ka Kui, Huang, and for the first time I knew that Jackie Chan could sing. At that time, I didn't have a penny to my name, and having an old-fashioned radio was a great happiness. A few years later, life conditions improved, and I could have a Walkman until the late 1990s when VCDs became popular. But I still liked Voice of Asia the most.⁵⁰

This series of messages, posted on Weibo by a former listener and reposted on Facebook by Wu Rui-wen, takes us from radio broadcasts through to the age of social media (see Figure 6 below for original post in Chinese). It demonstrates the persistence of cross-strait communication throughout shifts to different communication mediums, and indicates that even with its final broadcast, Voice of Asia did not end. This is because Voice of Asia was not just about the content of the broadcasts. Instead, it was about the long-lasting relationships formed through these broadcasts.

⁴⁹ Interview with author, March 10 2023

⁵⁰ These are selections from the longer original series of messages. From <https://www.facebook.com/lidaminging>, accessed March 2023.



Figure 6. Weibo messages sent to Wu Rui-wen from a former listener, which Wu Rui-wen then reposted on Facebook on April 26, 2023. See above for English translation.

The golden era of RTI marked a decentralization and de-formalization of the media, representative of changes taking place across the globe at the time. With broadcast radio conceptualized as multimedia and multidirectional, Voice of Asia became an important site for interpersonal, cross-strait exchanges, which were unprecedented prior to this period. These exchanges shifted how people conceptualized cross-strait relations and how they imagined themselves and their own ideologies in relation to others. In the late 1990s, declining cross-strait relations and the emergence of the internet led to a decrease in the popularity of radio broadcasts. Moving forward, the internet enabled on a small scale the long-term persistence of the relationships formed during this time.

The golden era of RTI can be conceptualized as establishing a model for how cross-strait relations and communication persists amidst cycles of political tension and the regulation of communication technologies. First, this period demonstrates the importance of direct exchange—whether through letters, phone calls, in-person visits, emails, or social media—in the shifting conceptions of the cross-strait socio-political imaginary, with these interpersonal exchanges complementing more official media narratives. Second, this period indicates that these exchanges can take place through the leveraging of latencies or gaps in the regulation of communication technologies. When regulation is nonexistent or insufficient, people find more opportunities to speak and listen to one another. Finally, the emphasis on the presence of exchanges and dialogue more so than the content of the dialogue is important here. Decades later, the content of the letters was difficult for listeners and broadcasters to recall, but the deep feelings of connection established through the very act of communication has remained strong and continued through to the present day, persisting across developments in communication technologies and changes in political relations.

3. Out of the Echo Chamber: Cross-Strait Exchange During the Clubhouse Moment

In February 2021, the social media app Clubhouse surged in popularity in the Mandarin-speaking world. As an audio-based social networking app, Clubhouse featured virtual chatrooms where its users could participate in live voice discussions on a wide range of topics with people from around the world. Clubhouse rooms could have anywhere from two to 5,000 participants, with larger rooms moderated by hosts and featuring tools like hand-raising for audience participation. Already a year into the COVID-19 pandemic, Clubhouse offered an antidote to the loneliness and isolation caused by lockdowns in much of the world at the time by providing the opportunity to listen in real time to other people's voices.⁵¹

On Clubhouse, people in Taiwan and mainland China found a rare opportunity to speak and listen directly to one another. This opportunity was exceptional in an environment with few possibilities for in-person exchange. After increases in cross-strait travel, business, and student exchanges in the 2010s during Taiwanese President Ma Ying-jeou's terms (2008-2016), governmental negotiations between the two sides ended in 2016 with the election of Taiwanese president Tsai Ing-wen (and mainland China's refusal to negotiate with her). Since then, mainland Chinese tourism to Taiwan declined and was fully suspended in 2019, even before the onset of COVID-19 restrictions in early 2020. The number of mainland Chinese exchange students in Taiwan peaked in 2015 and declined through 2020, when these students were fully barred from entry to Taiwan (C. Y. Lin and Lee 2020).⁵² While announcements from both governments credited COVID-19 for the suspension of exchange student programs, the Chinese Communist Party (CCP) reported that they also took into account the current state of cross-strait relations (Lo 2020). These policy changes were accompanied by increasingly cynical attitudes toward mainland China on the part of the Taiwanese public. Negative reactions toward Beijing's national security law in Hong Kong combined with a decrease in the economic benefits of cross-strait relations with the onset of COVID-19 restrictions in 2020 meant that mainland China's image and reputation in Taiwan "sunk to new lows" (Rigger 2021, 194).

Growing political tensions were also felt online, where Clubhouse became an exception to increasingly isolated internet environments. In mainland China, the censorship and surveillance mechanisms of the Great Firewall have maintained a high degree of control over the type of information that citizens can access since the early days of the internet, and especially since Xi Jinping became general secretary in November 2012 (G. Yang 2022, 19). The CCP's internet regulation efforts are supported by a citizen army of paid commentators (五毛), as well as young, nationalistic internet users often referred to as "little pinks" (小粉紅) who aggressively spout government propaganda on both mainland Chinese and international platforms.⁵³ Coupled

⁵¹ At the time, much of the world (including many parts of mainland China) was experiencing lockdowns. Taiwan was the exception, largely avoiding COVID-19 lockdowns through May 2021 (Wei 2022).

⁵² Though mainland Chinese students were allowed to enter Taiwan as the pandemic subsided, the number of these students in Taiwan remains significantly lower than pre-2020, with 25,049 students in Taiwan in 2019 and only 4,651 in 2023 (Mainland Affairs Council).

⁵³ The paid commenters are often referred to as "wumao" (五毛), which translates to "50 cents" in reference to the amount of money these commenters are paid by the mainland Chinese government per post. For detailed studies of this phenomenon, see (King et al. 2017) and (Han 2018). Guobin Yang writes that the term "little pinks" indicates "a fainter version of the Red Guards" and that these nationalistic internet patriots first appeared in 2016 with a cyberattack on newly elected Taiwanese president Tsai Ing-wen's Facebook page (2022, 183).

with mainland China's homegrown economy of propriety online platforms, the internet in mainland China looks very different than it does for users in Taiwan.⁵⁴ While internet users in Taiwan (and much of the world) are accustomed to using Google as their search engine and Facebook for their daily interactions, users in mainland China use Baidu and Weibo. This difference in platforms creates a wide gap in the information that is readily available on each side of the Taiwan Strait. Taiwanese people and the government are also often wary of mainland Chinese platforms, and the use of these platforms in Taiwan has at times become a geopolitical issue. For example, in April 2020 the Taiwanese government, citing concerns that the widely used videocall app Zoom's data was being routed through mainland China, banned the use of Zoom by government agencies in Taiwan (*BBC News* 2020).⁵⁵

It was in this environment of increased political tension, fewer opportunities for in-person exchange, and fractured internet communication that Clubhouse briefly emerged as a rare online site for open communication across the Taiwan Strait. During this time, Taiwanese journalist Liya Chen published an article with the headline, "If there was no Firewall, what would we talk about with each other?" (如果沒有牆·我們會跟彼此聊什麼?) (2021). The "Clubhouse moment" provided a brief window into a world without the Great Firewall, where individuals in Taiwan and mainland China could openly discuss their ideas and opinions online.

The Clubhouse moment was something of an anomaly not only in the tense environment of cross-strait relations, but also in the realm of social media today. Indeed, interactions on social media are often discussed as existing within "echo chambers," where internet users are exposed only to ideas similar to their own (Terren and Borge 2021; McPherson et al. 2001). In contrast, on Clubhouse internet users in Taiwan and mainland China briefly found the opportunity to confront diverging opinions. While online echo chambers are often thought to capitalize on human tendencies to interact only with like-minded individuals, the Clubhouse moment challenged this idea by revealing a willingness and excitement to confront differences through speaking and listening.

This chapter explores the conditions that led to Clubhouse's brief popularity (or "virality") as well as the specific affordances of live-voice communication on Clubhouse that created an online environment suitable for open exchange. In telling the story of the Clubhouse moment, I rely mainly on the perspectives of Taiwanese and mainland Chinese Clubhouse users who were active on Clubhouse surrounding the app's brief peak in early February 2021. At the time, Clubhouse required an invitation from a current user to join the app. By the time I found out about Clubhouse's popularity in the Mandarin-speaking world and managed to get an invitation, the hype surrounding Clubhouse was already declining. There was also no built-in way to record Clubhouse discussions at the time, and using a separate device to record was discouraged by the app developers and the Clubhouse community. Therefore, this analysis of the Clubhouse moment centers less on the interactions themselves and more on the discourse surrounding the event

⁵⁴ Guobin Yang discusses "platformization," or the idea that "social media platforms can shape user behavior through technological affordances in ways consistent with ideological norms." In mainland China, the heavy involvement of the CCP in platformization "is a process of the party extending its influence into social media platforms and apps" and therefore extending its control over the internet (G. Yang 2022, 36–37; de Kloet et al. 2019).

⁵⁵ Taiwanese people often choose to avoid mainland Chinese platforms, though people with connections to mainland China tend to use WeChat. Among the younger population of Taiwan, there is also growing interest in mainland Chinese social media apps like Douyin (the Chinese version of TikTok) and Xiaohongshu (小红书) (Graham-Harrison and Lin 2023).

through analysis of published articles, podcast episodes, and interviews with Clubhouse users. Specifically, I interviewed eight Taiwanese and eight mainland Chinese users in the months following the Clubhouse moment, either in person or online. Most of the people interviewed hosted chatrooms and were heavily involved on Clubhouse during its peak. I also engaged in informal discussions about Clubhouse with friends and colleagues leading up to and during my time in Taiwan. I then compiled these perspectives into a common narrative, drawing out themes and tensions across these various accounts.

The Story of the Clubhouse Moment

Clubhouse was first released in March 2020. As a Silicon Valley startup, the app slowly gained ground mainly among techies using the platform to promote their business ventures or discuss the latest cryptocurrency investment strategies (Griffith and Lorenz 2020). A key point in the app's popularization came on February 1, 2021, when Elon Musk appeared in a Clubhouse room that quickly reached Clubhouse's limit of 5,000 people (Newton 2021).



Figure 7. The Clubhouse logo. Reproduced from blog.clubhouse.com.

At the time, Clubhouse's slogan was "drop-in audio." Different from text and image based social media platforms such as Facebook and Instagram, Clubhouse features public and private group chatrooms in which people interact through live voice communication. While each user has a visual profile featuring a name, image, and any biographical information they chose to share, discussions on Clubhouse at the time could occur only through voice.⁵⁶ Clubhouse's early marketing positioned it as exclusive, initially accessible only on Apple iOS devices and requiring new users to receive an invitation from a current user. Once invited, users could join public chatrooms aligned with their interests or host their own public or private chatrooms. In large chatrooms, the host often selected a small group of people to participate in a speakers' panel, and only those individuals would have the ability to unmute themselves to speak. Everyone else in the room would listen to the discussion or raise a virtual hand to wait in line for the host to unmute them—in the largest rooms, this could take hours. When someone started to speak, a flashing gray circle would appear around their profile photo, visually linking their icon with their voice.

⁵⁶ Though Clubhouse users peaked in February 2021, the app is still available today. Since the early days of Clubhouse, new features have been added: hosts can record discussions and repost them later, and there is also a text chat feature for participants to type their thoughts.

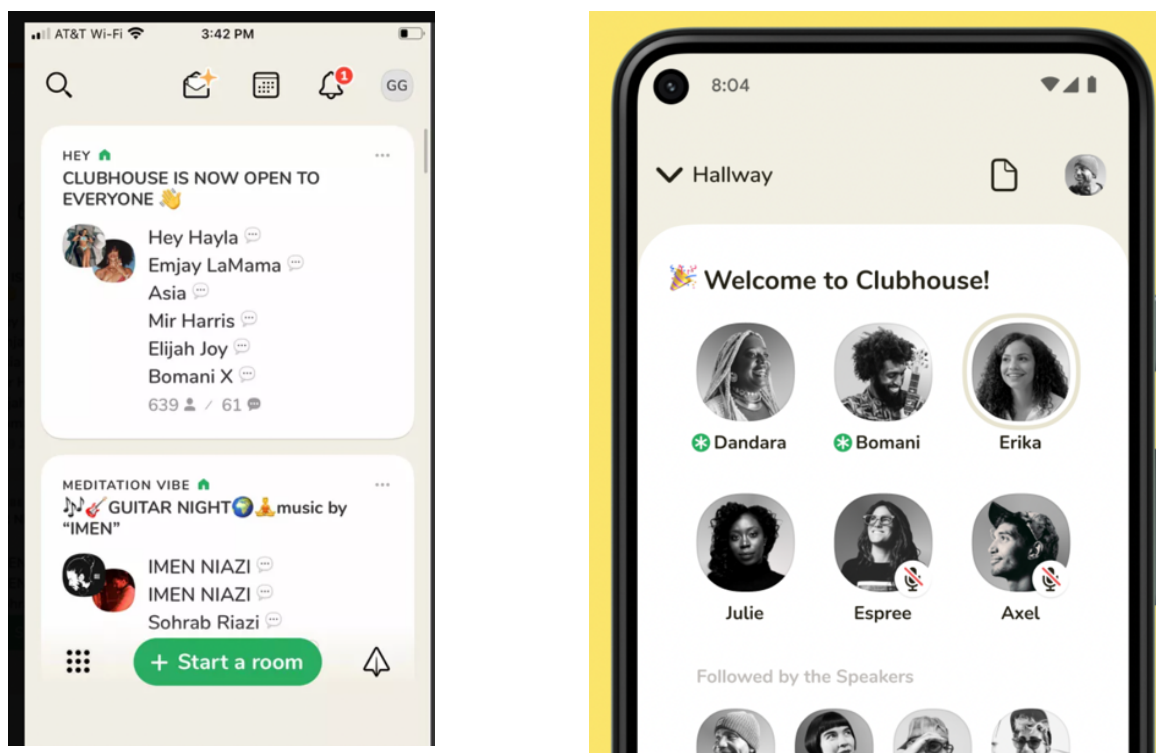


Figure 8. On the left, an example of a Clubhouse homepage where Clubhouse users can select a room on a topic that interests them. On the right, an example of a Clubhouse chatroom. The gray circle around the profile photo (top right) indicates the current speaker. The six users at the top are on the speaker panel. The four users at the bottom (cut off) are listeners and can raise their virtual hands to participate. Reproduced from blog.clubhouse.com.

Journalist Liya Chen joined Clubhouse slightly before its surge, in late January 2021. As a Taiwanese reporter and podcast host who covers cross-strait issues and lived in mainland China for three years, Liya was initially intrigued by Clubhouse's live-voice format. However, when she first signed on there were not many Mandarin chatrooms, so she joined some Japanese rooms to experience the app even though she could not understand Japanese very well. Soon after, Liya noticed more Mandarin chatrooms, and within a week or two it seemed that everyone she knew was on Clubhouse. At first, people opened rooms to talk about Clubhouse itself: how to use it or what it could be used for. Then some celebrities hosted rooms, and everyone was excited for the chance to speak live with their idols.⁵⁷ A key turning point for the Mandarin-speaking Clubhouse world took place on the evening of February 4, 2021, when the hosts of the popular Taiwanese news parody podcast *Bailingguo* (百靈果) opened a room in which they invited people facing oppression in Xinjiang and Tibet to talk to an audience of more than 1,000 podcast fans and their friends.⁵⁸ The room lasted for more than 12 hours. In a reaction video posted on YouTube the

⁵⁷ Interview with author, May 5 2022

⁵⁸ More than 1 million members of the Uyghur ethnic group have been arbitrarily detained in reeducation camps in Xinjiang by the Chinese government since 2017, and many more in the region are subjected to "intense surveillance, religious restrictions, forced labor, and forced sterilizations" (Maizland 2022). *Bailingguo News* has almost 900,000 subscribers on YouTube, and their most popular podcast video recordings can exceed 1 million views.

next morning, Bailingguo hosts Kylie and Ken discussed how rare it was to have the opportunity to speak directly with local people from Xinjiang and Tibet. They noted that this experience contrasted their typical online interactions with mainland Chinese little pinks who tend to be aggressive or defensive of the CCP.⁵⁹

This Clubhouse room seemed to open the floodgates to a topic that is rarely discussed openly in mainland China. Similar rooms began to appear in which thousands of Clubhouse users could hear directly from Uyghur participants. Reporter Jiayang Fan emphasized the rarity of this situation in a podcast interview with the *New Yorker* (Remnick 2021):

The conversations often circle to pretty taboo subjects, and those topics are not discussed because people don't like to leave a trail. So, hearing in real-time Chinese youths discussing subjects that they have never been allowed to discuss in classrooms or even really in cafes, that's incredibly intriguing and keeps me on there for hours at a time.

Also in early February, Liya noticed a surge of rooms focusing on the topic of cross-strait dialogue, many of which specifically encouraged young people from Taiwan and mainland China to discuss their political views. Given her expertise in cross-strait reporting, she decided to start her own chatroom. The room was titled, “Do you think there is real exchange happening in the rooms for cross-strait youth exchange?” (在兩岸青年交流的房間你覺得真的有交流嗎?), with the goal of creating a space for Clubhouse users to respond to and analyze the past few days of cross-strait discussions. By this point, however, the surge in Mandarin speaking users discussing political issues had caught the attention of censors, and Clubhouse was banned in mainland China on February 8—the same day that Liya opened her room (Goh 2021). The effects of the ban were not felt immediately, however. Since many mainland Chinese Clubhouse users were already accustomed to using VPNs to “climb the firewall” (翻牆), Liya’s room quickly attracted 1,000 people.⁶⁰ After two hours Liya had to leave, so she passed on the hosting responsibilities to someone else. When she came back to Clubhouse 24 hours later the room was still going strong, lasting for a total of 30 hours.⁶¹

While Clubhouse rooms focused on political discussion garnered global media attention and attracted large audiences, it is important to note that the platform’s appeal extended beyond politics. Many users joined to forge connections with like-minded individuals based on shared interests. One mainland Chinese Clubhouse user described in an interview how she met new friends from Taiwan through Clubhouse because of their shared interests in language learning and singing:

Since I joined Clubhouse in late February, most of the cross-strait rooms had already been shut down. But I could still find some interesting rooms related to the cross-strait rooms. Not about politics, but about cultures and languages. Some of the Taiwanese people were teaching the mainland Chinese people to speak *Minnanhua* (閩南話), and also

⁵⁹ 感想：在#Clubhouse 上開了跟新疆還有西藏朋友的聊天室. Bailingguo News. February 5, 2021. <https://www.youtube.com/watch?v=aLOW51Z3XMI&t=85s>.

⁶⁰ A VPN, or Virtual Private Network, is the tool that people in mainland China use to “climb the Firewall” (翻牆) by setting the location of their IP address to a different country in order to gain access to the non-Chinese internet.

⁶¹ Interview with author, May 20 2022.

some of the [Taiwanese people] are like ‘Oh, we want to learn *Beijing-qiang* (北京腔).’⁶² The mainland Chinese are from the northern part of China, and they are teaching the Taiwanese people to speak *Beijing-qiang*. And also some of the Taiwanese are teaching mainland Chinese to speak *Taiwan-qiang*. It’s really interesting to me, since I’m a language enthusiast and language learner, and I can speak both *Beijing-qiang* and *Taiwan-qiang*, so I did a lot of rooms about that.

I also met a lot of friends from Taiwan on Clubhouse, not because of politics, but because of music. I joined a club about singing and most of the [members] were from Taiwan, and we just sang a lot of songs together. I [was an exchange student] in [the Taiwanese city of] Taichung when I was in my year three [of college], so naturally I can set up a good relationship with the Taiwanese people. I’ve also...helped to host a singing competition. I think this is another kind of communication between Taiwan and mainland China. Not only Taiwanese people compete in the singing competition, but also some people from mainland China.⁶³

Clubhouse users I spoke with often discussed the depth of connections they formed with other users, often through shared mutual interests as in the chatrooms mentioned above. In some cases, connections that started online eventually progressed to offline friendships. I even spoke with a mainland Chinese couple that met on Clubhouse and were married a year later, and another Clubhouse user who met his boyfriend through Clubhouse.

After February 8, the effects of Clubhouse’s ban in mainland China materialized gradually. While some mainland Chinese Clubhouse users continued to use a VPN to access Clubhouse, others moved on to more easily accessible media. Some people also began to lose interest in Clubhouse as the novelty of the live-voice format wore off. Clubhouse users also discuss a qualitative shift in the content of Mandarin-language discussions that happened around the time that Clubhouse was banned. One mainland Chinese Clubhouse user speculated that the reason for this shift was the increased presence of little pinks on Clubhouse:

I think during the period that the government shut down Clubhouse, many [little pinks] flocked into Clubhouse because of some very famous KOLs such as Tuzhuxi.⁶⁴ He’s a very famous KOL that supports the Chinese government, and he has a Clubhouse account and opened many rooms to discuss and praise the Chinese government, and to say that people on Clubhouse talking about the issues in Taiwan or about Xinjiang or Tibet are nonsense... He had been on clubhouse for some days, then later the app was

⁶² “Minnanhua” refers to the Southern Min language, with variations spoken by people in Taiwan, Fujien, and parts of Southeast Asia. This is different from “Taiwan-qiang,” which refers to Taiwan-specific pronunciation of Mandarin Chinese. In this quote, she uses the Mandarin word *qiang* (腔) to refer to the Mandarin variations specific to the regions of Beijing and Taiwan.

⁶³ Interview with author, June 16 2022.

⁶⁴ KOL stands for Key Opinion Leader, which in this context refers to an online influencer that shapes their followers’ opinions on a particular subject (in this case, politics). Tuzhuxi, whose name translates as “Chairman Rabbit,” gained a following in mainland China in 2019 during the Hong Kong protests by sharing views online that tended to favor the mainland Chinese government perspective (C. Chen 2020). He is known for using his Western education as a way to bolster his credibility. In addition to starting his own rooms on Clubhouse (and bringing along his followers), one mainland Chinese Clubhouse user said that there were also chatrooms where users sarcastically praised him as a means of criticism (interview with author, January 30 2023).

suddenly shut down by the government, so in my personal opinion both things have some connections.⁶⁵

This account indicates that even though people continued to use the app after it was banned in mainland China, the quality of discussion had already begun to deteriorate. Clubhouse rooms became less about sharing and listening to diverse opinions and more about repeating deeply ingrained, unmoving beliefs. This decline in the depth of exchange can be thought to mark the end of the Clubhouse moment. While short-lived, the Clubhouse moment briefly provided a rare space for people to share ideas and connect across the Taiwan Strait, even amidst a broader context of increased communication restrictions and political tensions.

Out of the Echo Chamber

The in-depth exchanges that took place during the Clubhouse moment were unusual not only in the context of cross-strait relations, but also in the environment of social media—a space where interactions often trend toward division and discord. The concept of the “echo chamber” (同溫層) was first applied toward social media in the early 2000s to refer to situations in which “social media users selectively engage with like-minded others and ideologically-aligned content, thus rarely being exposed to the conflicting ideas that make up the agonistic public sphere” (Terren and Borge 2021, 100).⁶⁶ In the early days of the internet, the pessimism surrounding the negative implications of echo chambers contrasted more optimistic views about the internet’s potential to enable diverse communication and advance freedom and democracy through widespread and easily accessible online knowledge.⁶⁷

Echo chambers have since been widely discussed for their political consequences, and specifically for contributing to increased polarization and the amplification of extreme content through algorithms that optimize for user engagement (Terren and Borge 2021, 100; McPherson et al. 2001). These online dynamics align with what have been suggested as broader human tendencies to interact only with people similar to oneself and to seek information that affirms one’s own beliefs (Sunstein 2017; Hung 2024).⁶⁸ Algorithmic recommendation on social media platforms leverages these supposed human tendencies by selecting content in line with users’ past behaviors. The idea is that prioritizing this content will hold internet users’ attention for

⁶⁵ Interview with author, June 16 2022.

⁶⁶ See also Mouffe 2005; Sunstein 2001; 2017.

⁶⁷ For early internet optimism, see Terren and Borge 2021, 100; Gimmler 2001; Papacharissi 2002; Rheingold 2003; Benkler 2006. In the context of mainland China, Guobin Yang’s 2009 analysis reflect the optimism of the time. He wrote, “Online activism is a microcosm of China’s new citizen activism, and it is one of its most vibrant currents. In this sense, online activism marks the expansion of a grassroots, citizen democracy. It is an unofficial democracy because the initiatives, both in thinking and action, come from citizens. The expansion is evident both in consciousness and practice. In consciousness, the major developments are the rising awareness of citizenship rights among the Chinese people and the changing views of power and authority” (2009, 223).

⁶⁸ Sunstein defines the concept of homophily as “a strong tendency to connect and bond with people who are like them. The tendency to homophily is dampened if people live within social architectures that expose them to diverse types of people in terms of perspectives, interests, and convictions. But with an architecture of control, birds of a feather can easily flock together” (Sunstein 2017, 1). Because recommendation algorithms can exacerbate homophilic tendencies, there have been various strategies proposed for countering the negative effects of echo chambers by disrupting homophily (Raoufi et al. 2024; Tommasel et al. 2021).

longer. These recommendation systems then result in what are often referred to as “filter bubbles,” which “limit the novelty and diversity of the content that users are exposed to, and which—instead of contributing to viewpoint diversity—lead to online clustering and polarization” (Pariser 2011; Terren and Borge 2021, 100).

An online environment defined by echo chambers, filter bubbles, and increased polarization has been shown to provide fertile ground for the spread of disinformation, or the often coordinated efforts through which false and/or deceptive information is disseminated and spread online (Diaz Ruiz and Nilsson 2023). Online echo chambers and the resulting environments of polarization and disinformation carry significant consequences in the context of the Taiwan Strait. In Taiwan, echo chambers can create “openings for manipulation by foreign actors,” particularly mainland China, with disinformation shown to be a growing issue especially in the context of foreign influence in Taiwan’s elections (Hung 2024; Shen 2021).⁶⁹

The Clubhouse moment was therefore an exception in the media environment of the Taiwan Strait, which so often is characterized by restriction, polarization, and disinformation. Drawing on Kate Lacey’s discussion of listening in the public sphere, if the standard across much of social media is to “listen in to familiar voices,” Clubhouse provided the brief opportunity for Taiwanese and mainland Chinese internet users to “listen out” to voices and opinions different from their own (Lacey 2023). Clubhouse users often expressed surprise at the opportunity to communicate online across national, ethnic, socioeconomic, and political divides. For example, cross-strait reporter Liya noticed that mainland Chinese Clubhouse users seemed to come from a variety of geographic areas, including from small towns where people most likely would not have had opportunities to visit Taiwan:

Clubhouse at that time was a unique situation, because after this door opened wide, a swarm of people rushed in. Of course there are also restrictions, because there is a threshold for people who know about Clubhouse. I mean, of course in China it’s universal for everyone to use phone apps, but knowing about Clubhouse is a whole other thing. Chinese people in rural areas, they might not know Twitter. Assuming they know how to cross the firewall, after crossing the firewall they have to know they can go on Twitter. So Clubhouse has a threshold, but at least it’s more open, so it doesn’t restrict everyone as much. If you just know about this place, you can go in. So many listeners and participants from China appeared, and among them were some people, for example, from smaller areas, including someone from a small village. He said his friend introduced [Clubhouse] to him, so that’s interesting because he doesn’t live in a big city and he didn’t receive college education...He’s just a farmworker. So his life experience is very interesting. But this kind of person can’t apply to study in Taiwan and is not allowed to travel. He’s not qualified, so I think this is a huge difference.”⁷⁰

[Clubhouse 那時候是更特別的狀況，因為他是這個門打開了之後，一窩蜂的人衝過來。當然他也有限制，是因為知道 Clubhouse 的人還是有一定的門檻。就是說，當然在中國來講，大家使用手機 APP 的方式非常普及，可是你要知道 Clubhouse 又是另一件事情。中國的農民，他們可能就不知道 Twitter，他雖然知

⁶⁹ See chapter 5 for more discussion of disinformation in Taiwan’s political environment.

⁷⁰ Interview with author, May 20 2022. While cross-strait travel has faced increased restrictions since 2016, Liya’s comments indicate that even prior to this time there were financial and regulatory challenges that were difficult for people outside of large cities to overcome, leading to limited opportunities for travel for these groups.

道，假設他知道可以翻牆，可是他翻牆之後，他要知道有 Twitter 可以去，這又是另外一件事情。所以 Clubhouse 他有個門檻，但至少它是比較開放的，所以我覺得就比較沒有限制各種。你只要知道這個地方就可以進來。所以那時候很多的在上面的中國的聽眾或是參與者會出現，有一些人是比如說，他是一些來自小地方的就是中國很小很小的農村裡面的人。他就說，我聽我朋友介紹這裡，很有趣，他一向就不是住在大城市的那種可能，大學生受教的就是或是什麼工作者，他就是個農民工，很有趣，他的生活經驗就很有趣，但這種人肯定不能夠申請來臺灣唸書，或者是肯定不能夠通過陸客的，不過什麼就是觀光團，他沒有這個資格，所以我覺得這是巨大的差異。]

For the Clubhouse user from a small village that Liya mentioned here, Clubhouse likely provided a rare opportunity to speak with and listen to people abroad when his life circumstances prevent him from traveling. That said, people from small villages in mainland China were relatively rare on the platform, with discourse surrounding the Clubhouse moment suggesting people mostly came from a privileged group of educated, urban internet users (Kuo 2021). These perceptions are further supported by the fact that at the time of the Clubhouse moment, the app was only available for iPhone—the most expensive smartphone option owned by at most 15 percent of people in mainland China.⁷¹ Even so, Liya’s account indicates that Clubhouse did at least partially reach a subset of people who would otherwise have limited opportunities to engage with media from beyond the Great Firewall. This wide reach then created a rare opportunity not only for these individuals, but also for people who had the unique opportunity to engage in discussion with people outside of their typical online communities.

The rooms specifically dedicated to cross-strait dialogue also demonstrate a space outside of the echo chamber, with the stated purpose of the chatroom being for participants to express diverging opinions. While these rooms provided a unique opportunity for exchange and an initial sense of novelty, some Clubhouse users eventually expressed frustration with their experiences in these rooms. One Taiwanese student, NG, said that while listening to a discussion about Taiwanese independence, it was difficult to bridge the gap between worldviews shaped by the contrasting Taiwanese and mainland Chinese education systems. Mainland Chinese people have been influenced by government propaganda emphasizing the idea of a unified China, with Taiwan as an inseparable part of Chinese territory. In contrast, in Taiwan the education system increasingly emphasizes Taiwan’s history separate from China, including the period of martial law under the KMT and the transition to democracy starting in the late 1980s. At the same time, older generations in Taiwan were also subjected to propaganda during the martial law period, during which Chinese nationalism and reunification with China were promoted under the KMT’s rule. The effects of these different educational systems were audible in the cross-strait Clubhouse rooms. NG said:

I remember that we talked about the education and the history, and the Kuomintang and the martial law period, and how that affects people who believe that we are part of China. And with this kind of discussion, I think it’s clear in talking about this to Chinese people that they are suffering from some kind of propaganda. Some of the generations in Taiwan are suffering from another kind of propaganda. And so at some point we are in the same boat—that we’re trying to find our own national identity or

⁷¹ The New York Times reports there are 215 million iPhones in mainland China as of March 2024 (Tobin et al. 2024).

ethnic identity. But then that discussion is irrational. It's just based on what you have seen in China and what we have seen here in Taiwan, and there is a gap. A lot of discussions are based on addressing this gap. So the conversation is very long because everybody is joining and sharing their own thoughts, so it's a bit unorganized.⁷²

On Clubhouse, NG could engage in direct conversation with people who hold different worldviews and opposing political opinions, accessing a diversity of ideas that is typically limited for internet users confined by echo chambers and filter bubbles. In this account, NG expresses a sense of empathy in the recognition that both Taiwanese and mainland Chinese people have been influenced by propaganda, but also a sense of frustration in the characterization of the conversation as “irrational.” This sense of frustration stems from confronting the “gap” in knowledge between himself and people in mainland China.

Similar to NG, mainland Chinese Clubhouse user CY also expressed mixed feelings regarding the cross-strait Clubhouse rooms. He said that while at first the cross-strait discussions were novel and interesting, they soon devolved into repetitive disagreement. He eventually stopped participating in the large cross-strait rooms and instead started a smaller study group for academic discussion related to cross-strait issues, which was active for months after Clubhouse was banned in mainland China. He recounted:

At least for me, at first, it's a very exciting room. We can hear different voices and different opinions from both sides about their understanding and about their imagination of cross-strait relations. But soon after, in my opinion, we noticed that people are just repeating...some ideological slogans from both sides without deep discussion...People quarreling about those slogans is just the like the big loudspeakers between Amoy and Quemoy in the 1950s.⁷³ It's very similar to that. People use the loudspeakers to shout out the slogans to the other side without wanting positive feedback. It's similar to a political mission but in a different way. Although I don't think the people in that room really have some political mission, but the situation is very similar and it's very boring and that's the reason why we escaped from that room and started to host our own groups to discuss different things.

This quote is especially interesting because it indicates the presence of the rhetoric of ideological warfare transmitted by individual social media users in Clubhouse rooms. While in the 1950s propaganda was spread through centralized loudspeaker systems (as discussed in Chapter 1), here the psychological warfare efforts have dispersed and decentralized, now perpetuated at the citizen level. This account also indicates that for the most part discussions on cross-strait issues involved sharing divergent opinions but stopped short from arriving at consensus or common ground. Instead, participants often tired quickly from the hard work of listening to people with different opinions.

⁷² Interview with author, February 22 2022.

⁷³ Here he is referencing our previous discussion that situates discussions on Clubhouse within a broader history of cross-strait audio communication, beginning with psychological warfare via loudspeakers in the 1950s-1980s (see Chapters 1-2 for more detail). “Amoy” refers to Xiamen and “Quemoy” refers to Kinmen, the sites where the loudspeakers are located.

The audio format of Clubhouse therefore provided a unique opportunity for an in-depth confrontation of divergent viewpoints—an occurrence that is notably scarce in the digital realm. Clubhouse users displayed an initial willingness and enthusiasm to engage with perspectives that differed from their own, challenging prevailing notions of homophily that emphasize human inclinations to interact with like-minded individuals. However, despite the initial enthusiasm, discussions on Clubhouse often failed to bridge the gaps between opposing views. This limitation can be attributed, at least in part, to the brevity of the Clubhouse platform and the deeply entrenched nature of political beliefs in this context.

A Cross-Strait Public Sphere

In the previous chapter, I discussed the formation of a shifting cross-strait socio-political imaginary through multimedia exchanges surrounding broadcast radio in the 1990s. The Clubhouse moment also can be considered to have contributed to this cross-strait sociopolitical imaginary. In both of these contexts, people in Taiwan and mainland China found the opportunity to speak and listen directly to each other outside of the bounds of government-enforced media restrictions, with the very act of communication playing as large a role in this sense of novelty as the content of the exchanges. The Clubhouse moment was unique, however, for the depth of exchanges and the profound engagement of participants. In addition to contributing to the cross-strait socio-political imaginary, the Clubhouse moment can be considered to indicate the brief formation of a cross-strait public sphere, with public sphere briefly defined as “a shared environment for information seeking, debate, and opinion formation” (Terren and Borge 2021, 101).⁷⁴ While discussions on Clubhouse may not have reached the level of extended debate, the Clubhouse moment did involve a shared environment where participants could seek information, at times leading them to question or change their own beliefs.

With echo chambers thought to impede spaces for public discourse, the Clubhouse moment in contrast led to a brief disruption of online echo chambers. Indeed, the Clubhouse moment diverged from the typical cross-strait internet environment where interactions are often brief, text-based, hindered by mainland China’s internet controls, and overwhelmed by trolls or little pinks propagating slogans without deep engagement. Instead, Clubhouse briefly existed as a space where people could develop and exchange diverse opinions, share their concerns, and seek information. In other words, the Clubhouse moment embodied the optimistic ideals of the early days of the internet and social media, where these tools held potential to advance freedom of information and democratic ideals.

According to media historian Kate Lacey, the media is central to the formation and conception of a public sphere. Lacey writes, “While the public sphere does extend beyond the media, it is the media—in all their variety and ubiquity—that have come to constitute the idea

⁷⁴ Building on Habermas’ *The Structural Formation of the Public Sphere* (Habermas 1991 [1962]), Dahlgren writes that the public sphere can be considered as “a constellation of communicative spaces in society that permit the circulation of information, ideas, debates, ideally in an unfettered manner, and also the formation of political will” (Dahlgren 2005, 148; Terren and Borge 2021, 101). Terren and Borge write, “Social media have the potential to be a free and autonomous space for information and communication among citizens, contributing to the public sphere as envisaged by Habermas (1991[1962]) and Dahlgren (2005). However, this potential is not realized when diversity is lacking, when there is no (or little) exchange of opinions, no reasoned debate between opponents, and therefore no common ground or shared concerns” (Terren and Borge 2021, 101).

and experience of publicness. It is in the mediated space where we most often and most persistently encounter ‘the other’” (Lacey 2023; Silverstone 2006). However, the media tends to amplify the voices of privileged or dominant groups, thus requiring marginalized or dispossessed groups (what American literary critic Michael Warner calls “counterpublics”) to search for ways to make themselves heard (Warner 2002). This process often involves learning “to speak in certain ways in order to be heard in public spaces that were set up to recognize and validate certain forms of speech over others” (Lacey 2023). Solutions to this “crisis of voice” are often expressed through the rhetoric of “giving people a voice,” which Lacey argues is limited within the liberal western framework that conceptualizes voice as property (Lacey 2023; Couldry 2008).⁷⁵

Lacey suggests that an alternative to a public sphere centered on voice might be a public sphere centered on specific intersubjective listening practices. She explains, “Being part of a listening public means listening alongside others, inhabiting a space of plurality and collectivity—an intersubjective positioning which goes against the grain of an ideology grounded in individualism and the idealization of voice.” Importantly, this type of listening is not inherently linked to a politics of voice, since “it is possible to listen without listening *to* anything.” She specifically suggests the practice of “listening out,” defined as a “radical openness” and “anticipatory disposition” which opens space for a variety of voices to exist through the creation of a receptive audience, thus welcoming a diversity of voices in the public sphere. The technique of “listening out” can be difficult and challenging, but in Lacey’s conception “it is an essential technique of democratic political life” (2023). If existing online always within an echo chamber is the easy and convenient choice, listening out provides an alternative option for those willing to put in the work.

The Clubhouse moment can then be understood as the brief formation of a cross-strait public sphere based less on the act of speaking and more on the act of listening. In a typical Clubhouse room, there is only one speaker, but there can be thousands of listeners. These listeners are silent yet visible through their icons on the screen (Figure 8). Here, the listening public is front and center, silent but nonetheless integral to the medium. If the typical cross-strait media environment either prevents open exchange or gives voice to official political rhetoric, Clubhouse’s emphasis on listening affords the opportunity for its users to “listen out.” During the Clubhouse moment, the cross-strait media environment briefly restructured for those involved. This environment was no longer conceptualized as a Taiwanese public in opposition to a mainland Chinese public. Instead, Clubhouse users from mainland China and Taiwan became part of a single public sphere (part of this exclusive “clubhouse”), willingly identifying themselves as belonging to this collective opportunity to imagine cross-strait relations less as defined by vexed national boundaries and more as defined by processes of listening to one another.⁷⁶

⁷⁵ Regarding the liberal Western framing of voice as property, Lacey writes, “The privileging of voice in liberal constructions of democracy went hand in hand with the rise of capitalism and the framing of freedom of expression as a property right. In this dominant Western imagining, voice as self-expression figures as a personal belonging or as a product to be circulated and exchanged in the free marketplace of ideas” (Lacey 2023).

⁷⁶ Here I am drawing on Michael Warner’s conceptions of publics, specifically that a public is formed “from the rhetoric through which is it imagined,” as well as the idea that people become part of (or excluded from) a public by nature of identifying with (or lack of identification with) a public address (Warner 2002, 50, 58).

It was perhaps the scale and significance of this shifting socio-political imaginary that meant that the Clubhouse moment could only be short-lived; indeed, thousands of people on a social media app were not powerful enough to counteract decades of deeply ingrained political and social divisions. That said, the Clubhouse moment demonstrates that social media communication is not by definition characterized by echo chambers and polarization; instead, this polarization is the result of specific design choices on social media. In fact, there are recent efforts to specifically engineer online spaces that optimize for interactive engagement. Two examples from Taiwan (which has a vested interest in combatting the disinformation from mainland China that increasingly echoes through social media platforms) include vTaiwan and Miin (迷音). The first, vTaiwan, was launched in 2014 by Taiwan's digital minister Audrey Tang and group of "civic hackers" during and following the Sunflower Student Movement protests. The app combines online and offline engagement to build consensus rather than polarization through an "attitudes map that shows users the relationship of their opinions to the opinions of others," giving visibility to spaces of agreement rather than polarization (OECD 2021; vTaiwan 2023). The second is a "secure information platform" called Miin, developed by Taiwan AI Labs with the goal to combat disinformation by optimizing for balanced information rather than user engagement (Tu 2024). Unlike vTaiwan and Miin, Clubhouse was not engineered for the purposes of combatting the effects of echo chambers. Instead, the app works to optimize engagement through personalized recommendations of chatrooms for each user, with these recommendations based on a user's past activity, a user's manually selected interests in the app, and the activities of friends in the app (Kakar 2022). That said, the Clubhouse moment demonstrates that the app was briefly used as a way for people to communicate outside of their echo chambers, and perhaps the spontaneous discovery of this possibility on a mainstream social media app (rather than one specifically engineered for "listening out") contributed to the brief but widespread use of the app.

A Viral Listening Platform

Though Clubhouse was not specifically engineered to prevent the formation of echo chambers, the app briefly functioned in this way by providing a space for open dialogue and the brief formation of a cross-strait public sphere. For the remainder of this chapter, I will discuss three possible factors that contributed to this environment: (1) the virality of the Clubhouse moment, (2) the platform's emphasis on listening, which increases engagement and provides a space for the formation of affective connections, and (3) Clubhouse users' perceptions of safety and privacy using voice communication.

A Viral Platform

Karine Nahon and Jeff Hemsley define virality as "a social information flow process where many people simultaneously forward a specific information item, over a short period of time, within their social networks, and where the message spreads beyond their own [social] networks to different, often distant networks, resulting in a sharp acceleration in the number of people who are exposed to the message" (2013, 16). Based on this definition, virality is most often discussed in the context of specific items of information. However, in discussing the

virality of the Clubhouse moment, I focus less on viral content and more on the virality of the Clubhouse platform as a whole. The Clubhouse moment was a viral moment because so many people joined Clubhouse within the span of just a few weeks. The information items that rapidly circulated were invitations to the Clubhouse app, which functioned not as single message but as a gateway to a novel opportunity for communication.

The brief period during which Clubhouse went viral led to an opening in the cycle of regulatory latency, or the cycle in which regulation of communication technologies often lag behind the widespread use of these technologies, creating opportunities for unregulated and open communication.⁷⁷ In fact, an analogous process occurred with the social media app Instagram, which was banned in mainland China in 2014 following a surge in use, and which likely caught the attention of censors due to its widespread use in the student protest movement in Hong Kong at the time (BBC News 2014). While Instagram remained popular globally after it was banned in mainland China, interest in Clubhouse dropped off over the course of a few months after the ban, making the widespread virality of the Clubhouse moment even more striking.

As stated above, a key characteristic of virality is that information can spread to distant social networks, thus posing a contrast to the phenomenon of echo chambers where information is contained within a single online community. This dynamic was certainly present on Clubhouse, with large chatrooms hosting hundreds or even thousands of people. In a room of this size, it is difficult to imagine that everyone would share the same social network. With varied networks participating in the same rooms, it is also more likely for diverse political perspectives to be represented. That so many of these users joined Clubhouse during the course of a few months (or even a few weeks) may have further contributed to the ability of its users to avoid echo chambers. Indeed, the condensed time scale of the Clubhouse moment likely inhibited the recommendation algorithms that result in echo chambers. This is because when someone first joins a social media app, the app has limited data about their interests and preferences. For example, if someone joins Clubhouse and the first chatroom they enter is about cross-strait politics, the app might not know which side of the cross-strait political divide they fall on. This means the algorithms might suggest more rooms with people discussing this issue from various perspectives, giving the Clubhouse user the opportunity to hear from people who do not always share their own opinions.

The Clubhouse moment demonstrates new ways in which virality can be mobilized as a powerful tool. While virality has been shown to be an important phenomenon for political activism, it is most often discussed in terms of viral content such as memes (Mina 2018). The Clubhouse moment, however, was an example of a platform itself going viral, briefly opening new possibilities for communication until the widespread use of the app caught the attention of censors. Clubhouse's virality also in many ways fits perfectly within the cycle of regulatory latency. The brief popularity of the app enabled people to take advantage of the time period before Clubhouse was banned in mainland China. Just when Clubhouse reached peak virality and Clubhouse users began to lose interest in this new communication technology, regulators caught on and banned the app. In this framing, the Clubhouse moment demonstrates the broader potential of virality to be leveraged in the cycle of regulatory latency, enabling the rapid spread of information during brief windows of opening, before regulation is developed and implemented.

⁷⁷ See Chapter 5 for a more detailed discussion of regulatory latency.

Affective Listening

In addition to the virality of the Clubhouse moment, the app's emphasis on live voice communication also contributed to the brief formation of an online cross-strait public sphere. While Clubhouse's tagline is "drop-in audio," the types of audio-related activities on the platform overwhelmingly trend toward listening rather than speaking; at any given time on the app, there are more people listening than speaking. While listening does not necessarily imply "listening out," the experiences of Clubhouse users from Taiwan and mainland China indicate that there was a sense of intersubjective, collective listening happening during the Clubhouse moment.

A recurring theme in interviews with Clubhouse users was the way in which the emphasis on live-voice communication increased engagement and receptivity. On Clubhouse, participants found it more difficult to ignore or disregard perspectives that they disagreed with than they might have on a platform with a text-based option. One Clubhouse user who hosted chatrooms about psychology said that whereas on a platform like YouTube or Facebook he could choose to ignore certain typed comments, on Clubhouse he was more likely to let people have their moment to speak.⁷⁸ He explained:

With typing comments, you could have multiple people expressing their opinions, and it also means that the host could select the comments to reply. Whereas in Clubhouse you are forced to, I mean, you could interrupt but you also don't want to appear rude... For example, I was talking about psychology and then I had people who chime in, for example psychiatrists or people with a psychiatry background who tend to think that medicine can do everything for mental disorders like depression or anxiety. Whereas we psychologists think the opposite. Then there could be an audience member saying, 'Oh actually therapy doesn't work, you always need to start with medication.' So this person's opinions still gets a chance to be expressed, whereas if it's just a comment on YouTube, as a host I could just choose to ignore it.⁷⁹

On Clubhouse, one person speaks at a time. This means that if someone wants to participate, they have to listen to the comments of everyone in line before them while waiting their turn to speak. From the host's perspective, as this comment indicates, it becomes important to listen and respond to each and every comment to avoid "appearing rude," thus creating an environment where people with different perspectives feel comfortable expressing their ideas. The medium of Clubhouse thus affords a certain degree of engagement, which is often lacking on sites such as Facebook where typed comments can simply be ignored.

While Clubhouse users were in part compelled to listen to other people's comments to avoid being rude or while waiting their turn to speak, many people also discussed heightened emotional connections that came from the experience of hearing other people's voices. Liya, the

⁷⁸ In another account, Liya Chen compared Clubhouse to Facebook and said that Clubhouse is more interactive. On Facebook, it is easy to write a comment that makes people angry. In this case, the other party does not need to respond, resulting in a one-way interaction. On Clubhouse, however, moderators can choose people with different views and have them talk directly to one another (interview with author, May 20 2022).

⁷⁹ Interview with author, June 17 2022.

cross-strait reporter, said that she could get a better sense of people's emotions on Clubhouse than she could with written formats:

The authenticity and intimacy of text and voice are not the same. For example, if I want to answer in Clubhouse now, if I give you a text answer, it may be very concise, and you may not feel the atmosphere and emotion of what I am talking about. But when I am speaking about it, you may also listen and say, 'Hey, how does her voice sound when facing these issues? So it also has an emotional aspect. I think with [written communication], it is possible to achieve this type of an exchange. But I think the feeling that voice brings is still irreplaceable.⁸⁰

[文字跟聲音的真實度跟親密度不太一樣，比如說，假設現在我要回答在 Clubhouse，這個如果我是給你文字的答案可能會很精簡，然後你可能也不會感覺到我在講這件事情的氣氛跟情緒。可是我在講這件事情的時候，你可能也會觀察我說，哎，他的聲音面對這件事情聽起來是怎樣？所以他還會帶著一些感性的東西上。所以我覺得你剛剛講的那個方式，他卻是有機會可以彌補，或是說達到這樣的交流。可是我覺得那種聲音的感覺還是難以取代的。]

To better think through the differences between text and voice communication, it might help to imagine parallel discussions with the same content taking place through speaking and listening (for example on Clubhouse or a phone call) and through text (for example in a Facebook chat or an email). Written language excels at conveying the meanings of the words. But for spoken language, the listener also has access to the speaker's tone of voice, accent, and pronunciation, as well as the delivery of the words in terms of speed, pauses, or vocal register. In spoken language, the meaning goes beyond the words' definitions. As ethnomusicologist Jocelyne Guilbault writes, "Words are indeed never merely words. They do not come alone. While they themselves perform and can be performative... words in our case are also performed in sounds and sight. Together, they are part and producer of a sensuous and mindful experience as much for the speaker as for the listener" (Guilbault 2019).⁸¹

Voice communication holds more nuanced information for the listener than text communication, but perhaps less information than a video format (which includes sound as well as visuals). That said, some Clubhouse users also expressed a preference for audio-only communication rather than video media. In podcast host Kylie's words:

For me, because I'm a podcaster, I believe in the power of voice. A lot of times on YouTube or TV you are watching people, and you feel like they're performing something

⁸⁰ Interview with author, May 20 2022.

⁸¹ Guilbault goes on to argue that the ways in which words are performed and spoken has long been closely linked with the study of affect and emotion. She writes that in the 1970s and 1980s, discussions surrounding the non-linguistic aspects of language "led to a paradigm shift going from the study of text to the study of performance in multiple fields, including anthropology, folklore, emergent popular music studies, and Africanist art history." This shift resulted in the publication of a variety of texts emphasizing "participation, dialogism, call-and-response, repeated phrases and the aesthetic of bodily action and agency in hailing participants and creating the conditions of possibility for various forms of emotional bonding." Guilbault writes that "this shift to the study of performance could be said to have provided the opening for what came to be known as the anthropology of affect and emotion" (Guilbault 2019, 181). Similarly, in studies of popular music there was emphasis on going beyond the verbal meaning of lyrics to consider the ways in which lyrics are performed, as well as the groove and interaction in relation to the lyrics (Frith 1988, 112; Guilbault 2019; 174, Keil 1994).

because they're doing all kinds of acts or gestures. And it's pretty distracting. But for hearing a voice it's different. Because people say that your ears are your way of doing defense, but if they can break through your ears then they're pretty much putting down your guard. So it's easier to trust someone who speaks to you with voice rather than seeing someone on TV or YouTube.⁸²

While this comment in part addresses the differences of audio versus video mediums, Kylie here also points toward the informal social environment of Clubhouse, which enabled people to communicate directly and engage deeply with each other. In a media environment characterized by disinformation, propaganda, and distrust, this deep engagement was a rare and valuable experience.

The value of listening directly to people whose experiences differ from one's own frequently came up in the context of the rooms addressing the persecution of Uyghur people in mainland China. In an interview with journalist David Remnick on the *New Yorker* podcast, one Clubhouse user described a chatroom entitled, "Is there a concentration camp in Xinjiang?" In this chatroom, thousands of Mandarin-speaking Clubhouse users heard directly from Uyghur participants during a conversation that lasted 12 hours:

I think starting at the beginning of the room, there was a lot more denial. Most of the time, if you're Han Chinese, when you're watching a video of a Uyghur talking about his or her experience or her family's experience in a concentration camp, there's always an English voiceover. And then that makes it so much easier to fall into denial because, 'Oh, it's the Western media. They are trying to portray China as some sort of a political devil type of situation.' In Clubhouse, these people are speaking Mandarin. That person sounds like your friend, your neighbor. You immediately recognize his Northwestern accent because you went to college with somebody from that region. The distance closed up so quickly; you can almost feel like a lot of people are immediately drawn in. The connection is built at that moment, after some of the personal stories were shared. I think that was one of the reasons why this was such a phenomenal experience, because it forced people into a long overdue sense of empathy. It almost dragged all 5,000 people over the hill of denial that you wouldn't be able to climb by yourself (Remnick 2021).

Kylie recounted her own experience in this room listening to the story of a Uyghur Clubhouse participant, as well as the reaction of a mainland Chinese person to his story:

There was one time I remember really clearly when one of the Uyghur people was saying he believed in the Chinese system. He tried to survive, he tried to prosper, and he graduated from a pretty good college...he moved to many other cities to try to get a job and then once he got into a pretty good job, the boss suddenly told him, 'Sorry but you have to go because of your ethnicity' and he was like 'Why? I wasn't doing anything weird, I am just a normal person, a Chinese person.' But his boss said it's too sensitive now. So he was sharing that story. He was so frustrated. He went back to Xinjiang and he didn't know what to do. He said he feels like he has no future. And then at that point

⁸² Interview with author, October 6 2022.

another Chinese participant went on and said, ‘We never knew that this is what you guys are suffering and I feel so sorry and I apologize for what you have suffered.’ And that was... even now I feel goosebumps. I feel like this is a moment that they can finally speak to each other and see each other as human beings instead of how a lot of Han Chinese people will probably think, ‘Oh you know those Uyghur people are just a bunch of terrorists.’⁸³

The sense of empathy and connection discussed in these quotes point toward the “sensuous and mindful experience” of listening (Guilbault 2019), or what anthropologist Jing Wang defines as “affective listening,” which emphasizes relationality, coexistence, togetherness, and intentional engagement (J. A. Wang 2016). In the emphasis on togetherness and relationality, affective listening also takes on a political significance. As Lacey suggests, while “freedom of speech is a right ascribed to the individual... ‘freedom of listening’ by contrast inheres in the space *between* individuals, and is concerned precisely with guaranteeing the context within which freedom of expression can operate not as speech, but as *communication*” (Lacey 2013, 9). Building on this concept, artist Brandon Labelle writes that listening can give “resonance to voices often unsuited and unhomed by the political economies of attention and mediation” (LaBelle 2018, 87). Through affective listening, Clubhouse users created an intersubjective space for communication, giving resonance to voices that often go unheard.

With its emphasis on relationality, affective listening is well-suited for counteracting the polarizing effects of online echo chambers. Outside of the context of listening, emotional mobilization has also been shown to be a powerful tool in online activism. As Guobin Yang explains in the context of mainland China, strong emotions such as sympathy and anger have been mobilized for voicing outrage and indignation, while humor, play, and satire are often used for protests of government authority (G. Yang 2018, 1946).⁸⁴ With listening on Clubhouse often discussed as encouraging a sense of empathy and engagement, and with voice communication well-suited to conveying nuanced meaning through subtle shifts in tone, affective listening online may be especially well-suited for political activism through emotional mobilization.

Perceptions of Safety with Voice Communication

A third factor in why people were able to talk so openly and empathize with one another on Clubhouse relates to certain perceptions associated with the relative safety of voice communication. Regardless of whether a platform is permitted or banned in mainland China, the risks of censorship and surveillance are never far from the minds of mainland Chinese internet users. That said, though the live-voice format of Clubhouse does not guarantee safety or

⁸³ Han Chinese are the largest ethnic group in China, whereas Uyghurs are recognized as an ethnic minority. Interview with author, October 6 2022.

⁸⁴ Yang writes that a decline of internet protest activities in mainland China since 2013 is in part due to censorship techniques that “demobilize online activism by attacking its emotional and allegedly irrational character,” meaning that the mainland Chinese government has in fact recognized the power of emotional mobilization in online activism and shifted their strategies to monitor it (2018, 1946).

anonymity, some Clubhouse users discussed a perceived sense of security with voice communication.⁸⁵ One mainland Chinese Clubhouse user said:

I think at least for those political rooms, if you show your face or leave some textual evidence of your opinions about the politics related to China, it will be very dangerous, especially for those living in China. So that's the reason why Clubhouse gives an opportunity to most of the Chinese users to voice their opinions about Chinese politics. And not only about cross-strait issues, but also about other things related to China. So I think that's why Clubhouse was very popular at that time when people were talking about politics in February 2021. I think it is irreplaceable in terms of the different format of the communication.⁸⁶

This account indicates that some Clubhouse users felt more comfortable with the voice-centered format. This increased comfort level then encouraged them to talk more openly about sensitive topics.

While voice communication may have felt relatively safe compared to video or text communication, mainland Chinese Clubhouse users also took precautions even before the app was banned. One Clubhouse user I spoke with said that to try to hide her identity on Clubhouse, she used an international phone number rather than her Chinese number to sign up for the account, used only her English name, used a photo of a pet cat as her profile photo, and made sure not to discuss information about her university or where she lives. She also avoided rooms where famous dissidents were present, since these rooms are more likely to be subjected to surveillance.⁸⁷

Especially in the context of the Xinjiang rooms, some participants felt that the opportunity to share their experiences outweighed the risks. Balingguo podcast host Kylie said that while hosting rooms she would often ask participants if they were sure they wanted to speak, worrying that sharing their experiences publicly might put them at risk:

Some of them are overseas, and I feel like maybe they are safer. They have to worry about their passports but comparatively they are safer. But there are people who are still in China, so I asked them, 'Are you okay? Do you really want to share the story? Is that going to put you at any risk?' And they said that they were too desperate to tell their story. For them it's like there's nothing to lose. It's so bad already. They would rather risk their livelihood to tell their story. So that moves me too. Especially for the person who got kicked out from his good job and moved back to Xinjiang. There's no hope for him so of course he wants to share his story. That means more hope for him.⁸⁸

This account indicates that while for some people the live-voice format of Clubhouse led to increased senses of safety and security, others felt that this platform provided such a unique and

⁸⁵ See Chapter 6 for a detailed discussion of internet users' perceptions of voice censorship and surveillance in dialogue with technical expertise from researchers working in audio AI.

⁸⁶ Interview with author, June 16 2022.

⁸⁷ Interview with author, June 16 2022.

⁸⁸ Interview with author, October 6 2022.

rare communication opportunity that the benefits of sharing their experiences outweighed the risks. It was likely the combination of all of these factors—the perception of decreased risks associated with voice communication in combination with this rare, large-scale opportunity for open voice communication—that encouraged so many people to take advantage of the brief opportunity to leave their echo chambers and openly share their experiences in the cross-strait public sphere.

After the Clubhouse Moment

In contrast to established ideas that people tend to gravitate toward others with beliefs similar to their own, the Clubhouse moment demonstrated a widespread interest in opportunities for open dialogue between people in mainland China and Taiwan despite deeply ingrained political differences. It further illustrates the brief formation of a cross-strait public sphere, where people gathered information and reconsidered their own beliefs. Furthermore, it points toward the pressing need for this type of dialogue in a context of increased communication restrictions and political tensions. So much of the magic of the Clubhouse moment, however, was that it was short-lived and spontaneous. People rushed to participate in this moment because they knew it was fleeting. They had the rare chance to be a part of this viral event and did not want to miss this opportunity.

Due to the spontaneity of the Clubhouse moment, it would be difficult to recreate this dynamic so as to intentionally develop an online, cross-strait public sphere. That said, this event provides insight toward the unique characteristics that can lead to even the short-lived possibility of much-needed open dialogue across the Taiwan Strait. The Clubhouse moment demonstrated that voice communication can be a powerful medium for helping people to feel secure in highly restricted online environments, as well as to increase engagement, empathy, and affective connection. It also demonstrates the importance of taking advantage of rare opportunities for dialogue when they do arise, in an environment where online spaces increasingly place people in polarized echo chambers. In this context, perhaps what is most necessary is an attitude of readiness: a preparedness to rise to the occasion at a moment's notice when a rare opportunity for dialogue presents itself; and not only to tune in, but also to listen out by calling up the intersubjective listening practices that create space for the sounds of diverse voices to resonate.

4. Negotiating Political Differences through Online Play

In June 2009, millions of mainland Chinese World of Warcraft (WoW) videogame players flooded Taiwan's servers, triggering long waits to log on, lags in gameplay, and an environment of frustration among gamers from both sides of the Taiwan Strait. This "cyber diaspora" was a reaction to increased regulation in mainland China's gaming industry, which caused delays in the approval and licensing of games (Lin and Sun 2011).⁸⁹ The controversy centered around the second WoW expansion pack, which American videogame developer Blizzard released in the US on November 13, 2008. In Taiwan, where WoW ranked among the top three most popular videogames, the expansion pack was quickly released five days after its release in the US (L. C. Chen 2009, 40). In mainland China, however, the General Administration of Press and Publications rejected the release, mainland Chinese licensing company The9 Limited lost their WoW contract, and the licensing was transferred to Chinese technology company NetEase (網易).⁹⁰ In order to transfer WoW to their servers, NetEase took the game offline for a month, and mainland China's WoW gamers temporarily lost access to their game. The physical proximity of Taiwan's servers provided a workaround, since they hosted the most up-to-date version of the game in close geographic proximity to mainland China. Taiwan's servers allowed mainland Chinese gamers to avoid delays that might result from playing on a distant server, resulting in a virtual "mass migration" of WoW gamers across the Taiwan Strait (Lin and Sun 2019, 49).⁹¹ It was not until August 31, 2010—19 months following Blizzard's initial release of the second expansion pack—that NetEase finally received approval to release the latest version of WoW in mainland China.⁹²

This event was significant because it led to "unprecedented numbers of daily interactions" between people from Taiwan and mainland China amidst a long-lasting official relationship "marked by limited contact and political tension" (H. Lin and Sun 2019, 50). Many of these interactions were strained exchanges in response to frustrating gaming experiences, with this frustration caused not only by overloaded servers but also by differences in gaming habits

⁸⁹ Examples of increased regulation causing delays include a 2007 policy where online game publishers were required to incorporate anti-addiction software into their products and a 2009 policy requiring censorship software to be pre-installed on PCs sold in China (Zhan and Chan 2012, 193). This policy was supposedly to block pornography and violence, but also resulted in censoring of political discussion. However, mandatory installation was postponed due to public outcry. Szablewicz writes that by 2009 mainland China had reached a "peak" in "internet addiction moral panic," driven first by controversy over illegal internet cafes and later by fears of internet addiction (2020, 55).

⁹⁰ The9 and NetEase, like many Chinese technology companies, are not government owned, but do operate within the Chinese government's regulatory framework, which can impose significant limitations to companies' operations.

⁹¹ There was, in fact, a precedent to the 2009 "mass migration." Delays in the release of WoW's first expansion pack in mainland China in 2007 led thousands of mainland Chinese players to jump to Taiwanese servers, causing gameplay issues similar to the 2009 event. However, in this case the expansion pack was released quickly and mainland Chinese gamers soon returned to their home servers (H. Lin and Sun 2011).

⁹² WoW Insider contributing writer Scott Andrews explains that once mainland Chinese gamers could finally access the updated version of WoW, they noted some bizarre creative decisions attributed to censorship reasons: "When players logged in to China's version of Northrend, the reason for the additional six-month delay became apparent. Censorship had once again run rampant. Undead dragons became blue drakes. Zombies went from shambling, rotting corpses to bodies with smooth gray skin. Skulls, even decorative stone and metal skulls on items and buildings, were removed. The death knight's skeletal flying mount became a spectral griffin. Blood once again changed colors, this time to green. Weirdest of all, abominations became straw men" (2014).

and etiquette on Taiwanese and mainland Chinese servers.⁹³ Sociologist Ho-lin Lin and computer scientist Chuen-tsai Sun note that oftentimes gamers attributed these in-game frustrations to national, ethnic, or political differences: “What is alarming is that players’ feelings toward the other group are not directed against the other ‘player’ group, but rather against the other ethnic/national group. One Taiwanese player put it this way: ‘Chinese are uncivilized! I know that from playing with them in games!’” (2019, 50).

This type of rhetoric indicates that the tension at play extends beyond World of Warcraft toward the broader context of cross-strait relations. Indeed, while “Chinese players made threatening comments about taking Taiwan by force,” Taiwanese players used the term “locust plague” (蝗蟲) to describe the influx of Chinese players (Davies 2022, 90; H. Lin and Sun 2019, 50). This term, first used as a racist slur to describe the waves of mainland Chinese tourists that descended on Hong Kong and Taiwan in the 2010s, references a widespread animosity between Taiwanese and mainland Chinese people that circulates largely in online spaces (Rowen 2023, 127; Garrett and Ho 2014, 352; Kadar et al. 2013).

In January 2023, history seemed to repeat itself when mainland Chinese gamers again lost access to World of Warcraft. This time, Blizzard and NetEase had a falling out and broke off their 14-year licensing partnership. Yet again, this was in an environment of increased regulation of videogames in mainland China, with various delays in the approval of new game licenses since 2018 and the implementation of gaming time limits for minors.⁹⁴ Mainland Chinese gamers once again flocked to Taiwanese servers, resulting in technical issues and arguments about cross-strait politics. In a discussion on a Taiwanese online videogame forum, one comment read:

How many Taiwanese players were forced to quit by the locusts back then... and you still want to speak up for the mainlanders? Although there are a few kind-hearted players, most of them are just a bunch of trashy mainland *wumao*.⁹⁵

[當年的蝗蟲逼迫多少台灣玩家退坑...還要幫大陸人說話? 雖然有少數的善良玩家 大絕多數都是一群垃圾雜碎的大陸五毛。]

⁹³ Lin and Sun explain one example of conflict caused by differences in gaming etiquette practices: “One example entails the dungeon team practice of collecting dropped equipment after killing a boss. In the past, the default option among Taiwanese players was ‘greed,’ meaning, ‘I would like to have that piece of equipment if no one needs it;’ this is perceived as a gesture of courtesy. However, the Chinese default is ‘need,’ which gives the player a higher priority. When this cultural practice brought about constant conflicts and finally manifested as a confrontation between ‘Taiwan identity’ versus ‘China identity,’ the informal ‘need-before-greed’ practice underwent a significant adjustment in later game releases” (H. Lin and Sun 2019, 50).

⁹⁴ Specifically, in March 2018 the restructuring of the State Administration of Radio and Television led to a temporary pause in granting new game licenses and made the process of getting licenses for international games more stringent. Since then, the freeze of new approvals has continued. In 2019, restrictions were implemented to limit the time that minors could play videogames to 90 minutes per day on weekdays and three hours on weekends. In 2021, this time was further restricted to one hour per day only on Fridays, Saturdays, and Sundays. Approval of new game licenses was once again paused from July 2021 through April 2022, which “negatively impacted domestic gaming revenue... and contributed to a noticeable reduction of the market growth rate in 2021” (Inwood 2022, 3).

⁹⁵ The term “wumao” (五毛) translates to “50 cents” and refers to online commenters who are paid by the mainland Chinese government to spread propaganda online. Comment posted by username Yue Hen (玥痕) on November 21, 2022 on [gamer.com.tw](https://forum.gamer.com.tw/C.php?bsn=5219&snA=670357) (<https://forum.gamer.com.tw/C.php?bsn=5219&snA=670357>).

While hostility is a common theme in the discourse surrounding Taiwanese and mainland Chinese WoW interactions, focusing only on hostility overlooks a broader significance of these “cyber diasporas” in the context of cross-strait relations. In an environment where policies since 2015 have led to fewer opportunities for travel and student exchange programs between Taiwan and mainland China, and where mainland China’s Great Firewall increasingly restricts and monitors its citizens’ access to information from abroad, large numbers of mainland Chinese gamers found a space to interact with Taiwanese people on Taiwan’s WoW servers.⁹⁶ In this case, stricter regulations of communication across various sectors led people to find new opportunities for exchange.

This chapter asserts that online videogames provide a rare space for exchange amidst ongoing political tensions, in which individuals in Taiwan and mainland China can grapple with cultural, political, and linguistic differences in real time in the context of online play. It situates online gaming as a “pore” in the “porous censorship” of the Great Firewall (Roberts 2017, 1), drawing attention away from restrictions and toward the messages and meanings that make it through the restrictions.⁹⁷ While many of the WoW interactions were marked by disagreement and hostility, here I also explore alternative videogame contexts where Taiwanese and mainland Chinese gamers have a chance to speak with each other in-depth and directly through live voice conversations. The medium of videogames is especially suited to these types of interactions, providing a relaxed environment and encouraging a high level of engagement, while also functioning as a legal gray zone in the context of mainland China’s highly regulated media environment.

I specifically consider the context of Battle Royale videogames, such as Apex Legends and PubG, where individual players or teams compete to be the last one standing as the border of the game world closes in around them, forcing them into ever closer proximity.⁹⁸ In contrast to the aggressive backdrop of this game and the widespread misconceptions of the association of violent videogames with real-world violent behavior, the Battle Royale context of the examples in this chapter demonstrate a relaxed environment for play and social interaction.⁹⁹ In the realm of cross-strait relations, these examples reveal virtual exchanges that often defy the social expectations of the physical world. Even amidst the escalation of real-world conflict in the

⁹⁶ Tsai Ing-wen’s presidency in Taiwan beginning in 2016 was a turning point in the rate of tourists and exchange students coming to Taiwan from mainland China, as the CCP disapproved of her election and implemented measures to demonstrate this disapproval. These included cutting off cross-strait communication channels following Tsai’s election, which led to a decline in mainland Chinese tourism to Taiwan prior to it being fully suspended by the Chinese government in 2019. The number of exchange students also decreased prior to being fully suspended in April 2020 (C. Y. Lin and Lee 2020; Lo 2020). See Chapter 3 for more detail.

⁹⁷ Though extensive, the Great Firewall is also “porous” (Roberts 2017). While the Great Firewall is effective at deterring the vast majority of mainland China’s internet users from accessing what the government deems as sensitive information, a minority of wealthy, educated, and politically motivated individuals can often find ways to access this censored information. Roberts argues that allowing limited information to slip through is strategic on the part of the mainland Chinese government, since this strategy avoids calling unwanted attention to the severity of their efforts. See Chapter 3 for more details on this topic.

⁹⁸ While Fortnite is perhaps the most well-known Battle Royale videogame in the United States, Apex Legends and PubG are more popular in the Chinese speaking world.

⁹⁹ Psychologists Patrick Markey and Christopher Ferguson argue in their book *Moral Combat: Why the War on Violent Videogames is Wrong* that fears of the correlation between violent videogames and real-world violence are overblown and not supported by conclusive evidence (Markey and Ferguson 2017). See also: Markey et al. 2015; DeCamp and Ferguson 2016.

Taiwan Strait, the setting of virtual war here provides a way for Taiwanese and mainland Chinese people to confront their differences and, at times, find common ground. Having entered the virtual game worlds of PubG or Apex Legends, Taiwanese and mainland Chinese players—who often hold opposing political views—find the opportunity to play temporarily for the same team.

In the context of fast-paced Battle Royale gameplay, live voice communication is an essential tool: while using one's hands for a variety of tasks (running, shooting, etc), there is no time to stop and type. Using one's voice is the best way to strategize with teammates, and when the action of the game slows down conversations can turn to topics other than the game: food or music, differences in slang words and movie titles in Taiwan versus mainland China, or even current events. The context of online gaming therefore provides a setting through which to explore how people in Taiwan and mainland China connect in a live voice setting, as well as to analyze the tension and disagreements that arise during these interactions.

Videogames in this sense become a space of “track two diplomacy,” or a framework for diplomatic activity involving dialogue between experts outside of an official government capacity. These individuals are likely “familiar with the positions of their respective governments, but are not necessarily expected to assume or defend those positions in the dialogue” (Staats et al. 2019). Track two diplomacy was originally considered to be limited to facilitated workshops among influential citizens who did not have an official governmental role. However, there is also space to expand this definition to include “soft track two activities” with “conflict resolution experts, influential citizens, academics, and journalists” (Federer 2021). In this work, I position online content creators such as videogame livestreamers as “influential citizens” who can facilitate the processes of track two diplomacy, with videogames providing a space for Taiwanese and mainland Chinese individuals negotiate their differences through online play outside of the bounds of formal governmental negotiation.

Recorded livestreams of videogame interactions are often posted for an audience on the video streaming sites YouTube and Bilibili, providing a way to analyze these interactions through repeated viewing and in discussion with other gamers. These various remediations, from live gameplay to livestreaming to online video, open multiple levels of discourse: the conversations between the gamers during live gameplay, comments posted by viewers during and after the livestream, and commentaries on these recordings by mainland Chinese and Taiwanese gamers during group listening sessions conducted for this research. This chapter therefore focuses on how individuals engage with videogames “beyond” the content of the game, while also employing the method of listening to how people listen (Eidsheim 2019, 27; Feld 2015).¹⁰⁰

This specific genre of Battle Royale livestream recordings can be found on Youtube and Bilibili by searching a somewhat unexpected phrase: “cross-strait eating chicken” (兩岸吃雞). This phrase stems from an expression that appears on the screen when someone wins PubG, which can be considered a rough equivalent of the English expression “winner winner chicken dinner.”¹⁰¹ Presented to seem like chance interactions, many of these videos are recorded by

¹⁰⁰ Here I build on Jie Li's engagement with cinema audiences in the context of film exhibition networks in mainland China from 1949 through the 1980s, where these “grassroots audiences engaged with cinema *beyond* the film” (Jie Li 2022, 199).

¹⁰¹ The Chinese phrase is 大吉大利 今晚吃雞.

livestreamers who set out to find a teammate from across the strait as a way to create engaging content for their channel. While not necessarily chance interactions, these interactions also do not appear scripted, and their content serves as a site to explore how Taiwanese and mainland Chinese individuals discuss and represent their differences through live-voice communication in the relaxed environment of online play.

Videogames as a Medium for Political Dialogue

“I have a question for you, older bro. Why won’t you all let Pelosi come over?” One minute into a videogame livestream posted on YouTube, Taiwanese Vtuber Li Ting (李聽) poses this question to his mainland Chinese teammate (Figure 9).¹⁰² This recorded livestream of the Battle Royale game PubG (絕地求生) is titled “Cross-strait chicken eating: If Pelosi comes to Taiwan will China attack Taiwan?”¹⁰³

In this video, posted July 28, 2022, Li Ting is referencing then US Speaker of the House Nancy Pelosi’s visit to Taiwan, which occurred a few days later on August 2. The visit sparked intense backlash from mainland China, which views official contact between the US and Taiwan as inconsistent with its claims to Taiwan and responded with military drills that crossed the median line of the Taiwan Strait (Mozur et al 2022; Smith 2022). In Taiwan, mainland China’s military actions were not taken lightly. Civilian defense training courses amassed waitlists of thousands of people, and by December, Taiwan’s president Tsai Ying-wen increased the compulsory period of military service from four months to one year (NPR 2022; Simons 2022). As tension was building prior to Pelosi’s visit, Li Ting and his anonymous mainland Chinese teammate found a way to discuss the event in the context of online play. Occupying opposing sides of a geopolitical stalemate, in the virtual world of PubG they temporarily played for the same team.

As Li Ting flies his virtual plane he repeats his question, this time emphasizing the visual parallels of the game to real world events: “Why won’t you all let Pelosi take a plane over to Taiwan?” A few days later, Pelosi’s plane would land in Taiwan despite mainland China’s opposition, welcomed by crowds of Taiwanese people at the airport even while others remained at home indifferent or in opposition to the US’s involvement in local affairs.

In the virtual world, Li Ting’s PubG avatar jumps out of his plane and flies headfirst down to the map. He sports white hair in pigtails, with one tip red and the other blue.¹⁰⁴ His Vtuber avatar narrates the game from the bottom right side of the screen, wearing fuzzy cat ears and light blue headphones and sporting a gender-neutral haircut (Figure 10). As Li Ting descends downward, the mainland Chinese teammate answers his question about Pelosi. Amused, almost laughing, he says, “That’s impossible, impossible.” The mainland Chinese gamer’s avatar appears on the screen, and he and Li Ting release their parachutes and descend to the map together (Figure 11).

¹⁰² A Vtuber, or virtual YouTuber, posts videos online by speaking through a virtual avatar. The trend originated in Japan in 2016 and rapidly gained popularity internationally (Lu et al. 2021). Li Ting’s Chinese name is 李聽, which translates to “Lee Listen.”

¹⁰³ 李聽。《兩岸吃雞 裴洛西來台灣中國回答台灣嗎》July 28, 2022.
https://www.youtube.com/watch?app=desktop&v=_Asq9gQ3qWA

¹⁰⁴ The hairstyle is likely a reference to Harley Quinn in the 2021 film *The Suicide Squad*.



Figure 9. Li Ting asks, "Why won't (China) let Pelosi come to Taiwan?" Screenshot from YouTube.



Figure 10. Li Ting's in-game PubG avatar flies down to the map. His Vtuber avatar is at bottom right. Screenshot from YouTube.



Figure 11. Li Ting and the mainland Chinese gamer descend down to the map together. Screenshot from YouTube: https://www.youtube.com/watch?app=desktop&v=_Asq9gQ3qWA

Li Ting continues to press his teammate on the question, and the teammate continues to avoid the question in various ways.

Li Ting: But your foreign ministry said the military should prepare. What does “prepare” mean?

Mainland Chinese gamer: “Prepare” means get ready to take a boat across

Li Ting: Ohhhh to protect Pelosi, right?

There is a pause in the dialogue, until Li Ting gives an exaggerated, confused “uhhh?”

Mainland Chinese gamer: Right, right.

Li Ting: (Laughing) I don’t believe you!

They continue in this way as Li Ting’s avatar—now clutching a weapon with pigtails covered in a helmet—runs across a beach, a dock, and through a grassy field. Then, after an extended pause, the mainland Chinese gamer says, “What do you all think about all this?”

Li Ting, surprised, responds, “In the end you do want to know!” He continues to share his thoughts on the topic as the avatars run up a hill and into an abandoned building. “If Americans want to go somewhere,” he says, “they should more or less be able to go. How could they not let people go!”



Figure 12. Li Ting says, “How could they not let people go!” Screenshot from YouTube:
https://www.youtube.com/watch?app=desktop&v=_Asq9gQ3qWA

This interaction is significant—even exceptional—in the context of cross-strait relations. Amidst growing political tensions and facilitated by the unique affordances of online videogames, the gamers speak openly about their political opinions. In the previous chapter, I discussed the affective properties of live voice communication, which enabled people to connect and empathize with one another in chatrooms on the social media app Clubhouse. While Clubhouse was popular for just a few weeks before it was blocked by the Great Firewall, cross-strait voice communication through online games has persisted since the invention of Voice over Internet Protocol (VoIP) in the early 2000s.¹⁰⁵ Videogames therefore provide a long-lasting platform for voice communication in an area where there are otherwise limited opportunities for interpersonal exchange.

In addition to providing opportunities for live-voice communication, the medium of videogames has been discussed as encouraging a high degree of engagement.¹⁰⁶ While reading a book or watching television involve passive engagement, videogames and the associated action

¹⁰⁵ Invented in 1995, VoIP allowed internet users to connect their computers to speakers and a microphone and call each other online. Though VoIP was slow to catch on among general consumers (especially prior to improvements in audio quality with the shift to broadband internet in the early 2000s), it was used immediately in PC videogames, where live voice communication could significantly improve team game strategies and social experience (Gibbs et al. 2006).

¹⁰⁶ For example, see Brendan Keogh’s discussion of multi-directional embodied engagement between the gamer and the game (2018).

of play require active participation. In the context of Taiwanese horror videogames depicting difficult moments of Taiwan's history, Chee Hann Wu explains:

In playing the games, players experience, witness, and reenact the past by making their own choices and determining the direction of the game's plot within the given framework...Players become active agents in creating personal connections with the past, which further allows that past to be reimagined and most importantly, remembered" (C.-H. Wu 2023).¹⁰⁷

This active engagement is not necessarily limited to the past: as Li Ting demonstrates, even something as simple as imagery of an airplane can be enough to relate to or even embody current events. Whether engaging with historical narratives, as in Wu's examples, or a constantly shrinking battleground, as in PubG, videogames require that players actively experience the game through multiple senses: vision, sound, and haptic touch. Whereas in an online chat room it might be easy to zone out or quit at a moment's notice, Li Ting and the anonymous Chinese gamer—having been randomly assigned to the same team—are expected to stay online, interacting with the game and each other until the end of the round.

Videogames can also encourage empathy. As Ian Bogost discusses in a 2015 analysis of the videogame *Between*, one of the many uses of videogames is relating to others, whether through competition, collaboration, socialization, solitude, or even disconnectedness (Bogost 2015, 65). At their core, videogames involve playing a role, or "putting ourselves in the shoes of someone else," within the constraints of a complex model world (Bogost 2011, 4). Role-playing within these complex model worlds can then facilitate the formation of complex relationships. This dynamic may be especially pronounced with the use of audio communication; studies have shown that voice communication in videogames can be used to build trust, strengthen relationships, and avoid conflict, and that the voice medium combined with text may be more effective in these areas than text alone (Williams et al. 2007b).

In stepping into someone else's shoes, gamers also have the opportunity to temporarily disconnect from their real-world identity. In the PubG example, Li Ting takes on the role of both his Vtuber avatar and his PubG avatar. As a Vtuber, he facilitates the conversation as a way to create content for his channel. As a PubG avatar, he plays the role of the mainland Chinese gamer's teammate. Within these roles and through the guise of their avatar alter-egos, the two gamers find a way to breach certain conversational barriers. The gamers are then temporarily freed from indexical connections between their bodies (and associated gender, national, and political identifications) and their voices.¹⁰⁸ This freedom may encourage them to engage in a political discussion that in the real-world context of mainland China's highly regulated media environment would be difficult or perhaps even dangerous if they appeared only as their physical world selves.

¹⁰⁷ The games Wu discusses are *Detention*, *Halfflight*, and *Pagui 1 and 2*.

¹⁰⁸ Musicologist Nina Eidsheim discusses indexical connections between bodies and voices through the concept of the "micropolitics of listening" (Eidsheim 2019, 33).

Digital Boundaries and a Legal Gray Zone

Through the medium of online videogames, Li Ting and the mainland Chinese gamer found a rare opportunity for open dialogue even amidst tensions in cross-strait relations and the restrictions of the Great Firewall. This interaction, however, was not a chance encounter; rather, it was enabled by the distinctive infrastructural conditions of the internet and gaming regulations in the Taiwan Strait.

The internet is at times romanticized as a place where geographic boundaries cease to exist, as a site for seamless communication with people across the world, or as a frontier for free-flowing information open to all. In fact, this type of “technological triumphalism” did not begin with the internet, but was present even with the emergence of transistor electronics in the 1960s. Chinese literature scholar Andrew Jones notes that while transistor radios and portable cassette players did open new avenues for sounds to circulate, they also “operated across, within, and in complex and often complicit relationships with existing social topologies” (2020, 7). With the internet, the Great Firewall is representative of the great lengths governments will go to in order to reign in the seemingly endless possibilities of online communication. It has been decades since mainland China began the process of “nailing Jello to the wall” (as former US President Bill Clinton famously stated in 2000), embarking on what was previously thought to be an impossible task of restricting the extent to which people in mainland China could communicate with the outside world (Webster 2019; Griffiths 2019). That said, while online videogames are subjected to many of mainland China’s internet regulations, they also continuously challenge political boundaries.

Perhaps surprisingly, the infrastructures of online videogames often adhere to boundaries similar to those of the physical world. Lin and Sun write that “digital game technologies allow individuals from all parts of the world to play across geographic and social boundaries,” but with constraints built into the software settings.¹⁰⁹ These constraints often take the form of server assignments. For example, in *WoW*, there is one server for mainland China and a separate server for Taiwan/Hong Kong/Macao, and these separate servers “can determine which groups of individuals can or cannot interact and form communities” (2019, 49).

Just as physical borders can be crossed through travel (whether legally or illegally), so too can these “digital world boundaries.” This is why, for example, mainland Chinese gamers were able to hop over to Taiwanese *WoW* servers when they faced delays in the release of expansion packs. This is also why, amidst cross-strait political tensions and restrictions imposed through mainland China’s Great Firewall, Li Ting and the mainland Chinese gamer were able to engage and empathize with one another while playing *PubG*. Though videogame infrastructures often work to maintain existing geopolitical borders by translating physical world boundaries to digital ones, gamers continue to disrupt and question these boundaries.

One reason this disruption of digital boundaries is possible is that many of the most popular videogames in mainland China occupy what mainland Chinese gamer MZ calls a “legal gray zone.”¹¹⁰ This is how the “gray zone” works: for a videogame to be legalized in mainland

¹⁰⁹ This is true for hardware as well. On gaming consoles, for example, “hardware chips can support territorial lockouts,” so that someone using a Taiwanese gaming console might not be able to play a Japanese version of a game (H. Lin and Sun 2019, 49).

¹¹⁰ Interview with author, March 1 2023.

China, a tech company such as Tencent (騰訊) or NetEase needs first to purchase the game's licensing and then request a version number from the government. However, many of the most popular games in mainland China do not have a government-issued version number. These games are instead circulated through unofficial means as players can access them through a widely used international online videogame distribution service called Steam. In the case of PubG, for example, the initial publication of the international version of the game in 2017 led to a rapid rise of Chinese-speaking Steam users who signed up for the Steam platform for the sole purpose of accessing PubG. Tencent did not release an official Chinese version of the game until February 2021 (Wilde 2017; Roth 2021).¹¹¹

Games on Steam that do not have a version number also will not have a server in mainland China. Because of this, gamers need to use what is called an “accelerator” (加速器) to connect to an outside server so that their connection speed is fast enough for a seamless gaming experience. MZ explained in the context of the Battle Royale game Apex Legends:

You need to use an [accelerator] to access Apex, or your game experience will not be very stable. Actually, Apex has a server in Hong Kong. So if you live in the southern part of China, maybe you don't need an accelerator to play. But if you live far away from Hong Kong and you want to have a stable connection to the home server, you have to use an [accelerator], and you have to pay. So all the Chinese players have the habit of paying for [the accelerator] every month, or it's hard for you to play any foreign games.¹¹²

The technology for an accelerator works the same as the technology for a Virtual Private Network (VPN). A VPN is the tool that people in mainland China use to “climb the Firewall” (翻牆) by setting the location of their IP address to a different country in order to gain access to the non-Chinese internet. Though VPNs are illegal in China, the same technology is sold by some of mainland China's biggest tech companies when marketed as an accelerator subscription. In fact, one of the most popular accelerators is developed by the Chinese company NetEase.¹¹³

One result of this legal “gray zone” is that games that are popular in Taiwan also become popular in mainland China, even if they have not gone through the entire extensive legalization process. Mainland Chinese and Taiwanese gamers may end up connecting to the same server to play against each other or on the same teams. In fact, given Taiwan's geographic proximity to mainland China, it is more likely that Taiwanese and mainland Chinese gamers will encounter each other than people from farther away. Even with an accelerator, a large geographic distance can cause connection delays that make gameplay frustrating, if not impossible. As MZ explained, since moving to Europe she has been unable to play videogames with friends in mainland China, because even with an accelerator the lag time was too long.¹¹⁴

¹¹¹ The PubG example above was posted in July 2022, after the mainland Chinese version of PubG was released. It is unclear in this example whether Li Ting is using a server in mainland China to play the Chinese version of PubG or whether the mainland Chinese gamer is using Steam to play the international version of PubG.

¹¹² Interview with author, March 1 2023. While she uses the example of Apex Legends, this process is the same for PubG.

¹¹³ MZ said that everyone she knows in mainland China uses NetEase's accelerator (interview with author, March 1 2023).

¹¹⁴ Interview with author, March 1 2023.

The term “gray zone” is at times used in the context of the Taiwan Strait to refer to tactics that mainland China uses in attempts to influence and intimidate Taiwan. Though there is no precise definition for gray zone activities, they “are coercive and aggressive but designed to stay below the threshold of triggering major conflict” (Stokes 2023; Mazarr 2015; Green et al. 2017). In January 2024 during the weeks leading up to Taiwan’s presidential election, a surge in mainland Chinese balloons flying over Taiwan led various news publications to report an increase in mainland China’s “gray zone” tactics (Buckley and Chien 2024). The videogame “gray zone” that MZ mentioned, however, is not a space of governmental or military tactics. Instead, it is a space that challenges the digital borders of the Great Firewall while still remaining below the threshold that would catch the attention of censors. This “gray zone” therefore operates outside the bounds of state-level actions, hosting informal, unmoderated interactions where Taiwanese and mainland Chinese gamers can grapple with their differences in the context of online play.

Listening to Listening and the Connotations of Cuteness

There are various layers to a recorded videogame livestream. There is the medium of the game, which establishes the constraints of a model world. There are the infrastructural affordances and boundaries that influence who can play on which servers. Then there is the content of the game: the activities of the gamers within the virtual game world. The final layer here, that I have yet to discuss, is the audience. Livestreaming and/or recording a videogame invites spectators who might be attracted to the stream to improve their own gaming technique or as a form of entertainment. Here I draw attention to this audience level of discourse by considering how Taiwanese and mainland Chinese gamers perceive and react to videogame livestreams during group viewing sessions. I therefore apply the concept of “listening to listening” as a way to engage in dialogue surrounding the voice, moving past assumptions of the voice as “essential, innate, and unmediated” and instead situating the voice within the specific historical, social, and political relations through which it is produced, mediated, and perceived (Eidsheim 2019, 27; Feld 2015).

On a video call with mainland Chinese gamer MZ, I pull up the video streaming website Bilibili and load a PUBG video that she had recommended in our prior email exchange. The title at the top of the page says “Eating chicken, I came across a Taiwanese milky-voiced young brother! His milky milky accent is actually so sweet!?” (吃鸡偶遇台湾奶音小哥哥！口音奶萌奶萌的居然如此甜美！?).¹¹⁵ An avatar wearing a tattered white tank top and a black utility belt, who appears to be the livestreamer, runs through the ground level of an under-construction building. A voice says, half laughing, as if the recording started mid-conversation, “It’s easy to hear you.” The livestreamer responds, “Your Taiwanese accent is easy to place” (Figure 13).

At this point, it becomes clear that the streamer is from mainland China and that he has encountered a gamer from Taiwan. The Taiwanese gamer does not immediately respond, and in an extended moment of silence the mainland Chinese streamer’s avatar runs out of the building through an outdoor zone filled with storage containers. A lone upward guitar slide emphasizes the pause in conversation. Then the Taiwanese player says, “Where are you from?”

¹¹⁵ [生命幻象. 吃鸡偶遇台湾奶音小哥哥！口音奶萌奶萌的居然如此甜美！?] Bilibili. December 7, 2018.



Figure 13. Screenshot from Bilibili. The Taiwanese player says, “It’s easy to hear you.” The mainland Chinese player says, “Your Taiwanese accent is easy to place.”



Figure 14. Screenshot from Bilibili. The *danmu*, or timed floating comments, are in white at the top of the screen.

This initial dialogue is a first attempt to place each other in pre-determined categories—an attempt to answer the “acousmatic question” of “who is this?”¹¹⁶ But before they can continue, MZ asks me to stop the video and points to the written comments floating across the top of the screen. She explains that these comments are called *danmu* (彈幕), which literally translates to “barrage,” as in the military context: the first character means “bullet” and the second character means “screen.” While videogame players shoot rounds of bullets, the audience of videogame livestreams instead shoots rounds of comments (Figure 14).¹¹⁷

Only 13 seconds into the video, a barrage of comments already cover the top of the screen: “So cute!” “Hahahahaha cute!” “Haha this little guy is too cute!” “Hahahahaha this milky sound is too much...My ears...” “This sound is too milky, if you’re an adult, I would probably beat you up.”¹¹⁸

As a non-native Chinese speaker, “cute” and “milky” are not obvious associations for me with the voice in the video. But later in a videocall with mainland Chinese gamer ZYS, she explains to me that she often hears these types of words used to describe the way that Taiwanese people speak:

I hear these adjectives quite often, especially used to describe Taiwanese girls. Then they will say, if they have an accent, they are cute, milky (奶), adorable, or they will say babyish (嗲).¹¹⁹ [我還蠻常聽說·就是用這個形容詞去形容·特別是就是臺灣的女生·然後他們會就是說會口音會很可愛·很很奶, 很萌·或者是說談會很嗲。]

While ZYS has heard these words more often for Taiwanese girls’ accents, the video in question involves two boys. And while the barrage of floating comments (which on the Chinese platform Bilibili are most likely to be from mainland Chinese viewers) seem to agree about the overwhelming cuteness of the Taiwanese “little brother’s” accent, I later watched this video with Taiwanese gamers SR and NG, who did not agree with this characterization:

SR: It said cute, but I think it’s just a normal accent. [他就說可愛·但是我覺得他的口音就是正常的口音。]

NG: Yeah his accent isn’t really cute. From my perspective, it’s not cute at all. [對啊他的口音沒有這麼可愛·對我來說·那個口音也不可愛。]¹²⁰

This discrepancy—where mainland Chinese commenters think the accent is cute but Taiwanese gamers think it is normal—stems from complex socio-linguistic and historical

¹¹⁶ The concept of the acousmatic question was first suggested by Pierre Schaeffer in the context of the emergence of audio recording technology, which enabled people to hear a sound without seeing who or what made the sound. Eidsheim builds on this concept to suggest that the question of “who is this” is a futile attempt to place people in a fixed identity, arguing that voice is not a precise identifier and that we keep asking the acousmatic question precisely because it is so difficult to answer (Eidsheim 2019, 2).

¹¹⁷ Chinese media scholar Jianqing Chen describes *danmu* as a violent, disruptive force in the traditional viewing experience. By using *danmu*, spectators visually interrupt the moving image, breaking the traditional barrier that prevents direct interaction with the screen (J. Chen 2023). For more on *danmu*, also see Jinying Li 2017.

¹¹⁸ Original Chinese of the *danmu* comments: 好可爱/哈哈哈哈哈可爱/哈哈这个小哥特别可爱!/哈哈哈哈哈这个奶音有点炸...耳朵.../这声音也太奶了吧·你要是成年了·我大概会揍你。

¹¹⁹ Milky (奶) can be translated as cute in a childish way.

¹²⁰ Interview with author, February 18 2023.

reasons. Though Taiwan and mainland China share a common official language of Mandarin, there are local variations in vocabulary, grammar, and pronunciation. Linguist Chun-yi Peng demonstrates that in mainland China the most common perceptions of the Taiwanese Mandarin accent stem from the media. He terms this media representation as “mediatized Taiwanese Mandarin” to distinguish representations of Taiwanese Mandarin on TV from the diverse ways that people in Taiwan actually speak.¹²¹ Peng explains that this specific Taiwanese speaking style became familiar to many in mainland China in the 1970s, gaining a reputation of prestige and cosmopolitanism from its associations with Taiwanese TV shows (2021, 75). Recently, however, perceptions of Taiwanese Mandarin have shifted among millennials in mainland China. Based on surveys and large-scale social media analysis, Peng shows that many people in mainland China will report that Taiwanese Mandarin sounds “non-standard” or “awkward.” Classifications of “innocent,” “cute,” “babyish,” or even “emasculated” are also common (2021, 66).

MZ’s perceptions of Taiwanese Mandarin seems to represent these shifts. She commented, “I think the Taiwan accent is really iconic. Among my mainland Chinese friends, sometimes we will even mimic to speak like a Taiwanese, to make a joke about it. It’s not a bad joke. We just think, ‘Wow, the way Taiwanese talk is so cute, and it would be funny.’”¹²² In her use of “iconic,” MZ alludes to the association of Taiwanese Mandarin with the entertainment industry, while the characterization of “cute” follows Peng’s discussion of more recent associations.

Recorded cross-strait videogame livestreams provide a site from which to extend Peng’s findings beyond official media toward the context of casual conversation, as well as to consider not only how mainland Chinese people perceive Taiwanese accents, but also how Taiwanese people react to these characterizations of Taiwanese Mandarin. Exploring audience reactions to these videos, in other words, provides an opportunity to consider how Taiwanese people listen to mainland Chinese listening, with these perspectives informing understandings of the complexities of cross-strait relations across multiple layers of discourse.

Accents that mainland Chinese people may identify to be cute do not always sound this way to Taiwanese people. Taiwanese gamer ARZ said that even when she is trash-talking in a game, mainland Chinese people will comment that her voice sounds gentle. Though she disagrees with this characterization, she said that she can empathize with their perspective because she is also listening to the mainland Chinese gamers’ voices and comparing them to her own:

Usually talking about these things, it’s that both sides have their mics on to speak, so I can also say their speech is harder, because maybe their pronunciation is more clanky. In fact, I can hear it from both sides. I can hear my own voice, and I hear their voices, so I can intuitively compare it...and I can actually understand why they think Taiwanese people speak more gently, because our articulation maybe isn’t as heavy. So it doesn’t really reach the point where I feel scared to speak or offended, because there are voices

¹²¹ For a detailed overview of the various political and historical factors influencing language in Taiwan, see Price 2019.

¹²² Interview with author, March 1 2023.

on both sides, so there's a comparison. So I say, 'Thanks, but for people here, it's pretty normal to talk like this, so it's just a difference of what you're used to.'¹²³

[通常會講到這些，就是兩邊都有開麥克風講話，所以我也是可以意識到說他們講話確實比較硬，因為可能發音比較鏗鏘。確實可以從兩邊聽到了，我聽到我自己的聲音，我也聽到他們的聲音，所以我可以很直觀的對比...我可以理解為什麼他們會覺得臺灣人講話方式比較溫柔，因為我們可能咬字沒有那麼重粒子。所以也沒有特別到時候會覺得很害羞，或者是被冒犯，這些其實是沒有，因為當時兩邊的聲音都有呈現出來，我覺得好像有一個客觀對比。所以就會跟他們說「喔，謝謝，不過其實我們這裡的人，大部分都是正常，都是這樣子講話，所以可能就是習慣性的差別吧」。

While ARZ takes it upon herself to patiently explain linguistic differences to mainland Chinese gamers, Taiwanese gamer NG at times feels offended by these perceptions. Regarding mainland Chinese comments characterizing the Taiwanese gamer's accent as cute, he said:

I feel like it's a bit of a microaggression. I mean I think this video seems fine, but sometimes when I communicate with Chinese people and then they say, 'Oh your Taiwanese accent, this or that,' it will make me uncomfortable... It's that now it seems like lots of Chinese people will think studying Taiwanese pronunciation is interesting. They will just randomly use this kind of cutesy speech. But my subjective feeling toward this kind of speech is that it's like they are trying to take advantage of Taiwan. I've seen some Chinese girls and they purposely study Taiwanese pronunciation to make themselves a bit more cute.¹²⁴

[我會有覺得有一點 microaggression 喔，就是我覺得這個影片好像還好，但是我有時候跟中國人交流，然後他們就會說，欸你臺灣腔怎樣，怎樣的時候我就會覺得有點不太舒服喔...就是好像現在很多中國人會覺得學臺灣腔是一件有趣的事情喔，就是好像覺得亂講話，這樣很可愛或什麼之類喔。但是對這個口音主觀的感覺就會覺得你有一點想要吃臺灣豆腐吧。我有看過有一些中國女生會故意學臺灣腔，然後讓他自己變的更可愛一點。]

Perhaps NG takes offense at the “cute” characterizations of the Taiwanese accent because of certain associations with this aesthetic category. Sianne Ngai writes that cuteness is “deeply associated with the infantile, the feminine, and the unthreatening,” and that “objects seem most cute when they seem sleepy, infirm, or disabled” (2012, 54–59). Intrinsically linked to an imbalance of power, “the cute object is as often intended to excite a consumer's sadistic desires for mastery and control as much as his or her desire to cuddle” (Ngai 2005, 815). The simultaneous associations of cuteness with both weakness and aggression can help clarify why characterizing an accent as “cute” might be perceived as a microaggression. This also explains remarks such as “this sound is too milky, if you're an adult, I would probably beat you up” (mentioned above), which conflates cuteness with aggression.

In the context of Taiwan's position in cross-strait relations, the associations of cuteness with the powerless and unthreatening have heightened meaning. Barred from participation in international organizations like the United Nations and the World Health Organization, Taiwan's

¹²³ Interview with author, March 9 2023.

¹²⁴ Interview with author, February 18 2023

international isolation is often interpreted as a weakness.¹²⁵ Peng maps this interpretation onto language, proposing that mainland China's increasing economic and political influence—as exemplified by the growth of the mainland Chinese television industry and the decline of the mainland Chinese market for Taiwanese TV shows—has resulted in a decrease in symbolic value of Taiwanese Mandarin while elevating the status of standardized Mandarin (普通話) (Peng 2021, 65).

However, studies of cuteness in the specific context of Taiwan complicate the extent to which this aesthetic can always be associated with a lack of power. In a study of the mobilization of the kawaii aesthetic in Chen Shui-bian's successful campaigns for Taiwan's presidency in 2000 and 2004, communications professor Hsin-yen Yang suggests that the aesthetic of cuteness can also be interpreted as subversive. The kawaii aesthetic, loosely differentiated from Western concepts of cuteness by its “quirkiness,” began in the 1970s in Japan as a way for young people to rebel against traditional Japanese culture. It was soon adopted for marketing purposes and spread internationally, but the connotations with disrupting traditional or dominant societal forces remained (Yang 2023, 87–88; Kinsella 1995). Yang explains, “While seemingly cheerful and soft, the kawaii style has been an expressive form of refusing to submit to the oppressive culture and an attempt, often through consumption, to demonstrate strength and unity among the common people” (2023, 90). Though “cuteness” as an aesthetic category has associations with powerlessness, it can also be mobilized strategically to give voice to the powerless. Pushing this further, Ngai shows that cute objects “often have a deverbalizing effect on the subjects who impose cuteness on them,” where, for example, babies or puppies might “solicit a response along the lines of a murmur or a coo.” This means that “the cute object shows its ability to infantilize the language of its infantilizer” (Ngai 2005, 827). Cuteness here not only gives voice to the powerless, but also draws the voice of the powerful down to its level.

By “listening to listening” in the context of recorded videogame livestreams, it becomes possible to access the various layers of complex power relations as understood through live-voice communication, with each vocalization pointing toward specific histories and identities to be interpreted differently by each listener. While a comment characterizing a Taiwanese accent as cute may be interpreted as an aggressive display of power, a mainland Chinese woman imitating a “cute” Taiwanese Mandarin accent may also be an attempt to gain voice and agency.

Sajiao Drag

I'm back at the beginning of Li Ting's cross-strait chicken eating livestream, this time watching while on a videocall with Taiwanese gamer ARZ. In the video, Li Tin mills about a waiting area before the game officially starts and strikes up a conversation with the anonymous mainland Chinese gamer as they meet for the first time.

“Heyyyy big brother,” says Li Ting in a nasally, almost whiny voice (Figure 15). On the other end of our videocall, ARZ laughs at the exaggerated tone of his voice.

¹²⁵ In a discussion of “paradiplomacy,” or diplomacy at the sub-state level, Sarah Newland writes that Taiwan counteracts international isolation by engaging in various diplomatic efforts such as sister city programs and driver's license reciprocity (2023). Driver's license reciprocity in this context refers to agreements between Taiwan and individual US states that allow the mutual recognition of driver's licenses to facilitate easier licensing processes.

A lower pitched male voice, the anonymous mainland Chinese gamer, says, “Male or female?” (Figure 16).

Li Ting responds, “Male, male,” lowering his vocal register considerably as if to ensure his mainland Chinese teammate.

The mainland Chinese gamer responds, “Ah? That voice is so nice.”

Li Ting’s avatar runs to the side of the road into a dirt patch and throws a punch at a random avatar. He says, “Where are you from, baby?”

The mainland Chinese gamer responds, “Jiangxi,” and then immediately asks, “Taiwanese accent?” (台灣腔).



Figure 15. Li Ting says, “Heyyyy big brother.” Screenshot from YouTube.



Figure 16. The mainland Chinese gamer says, “Male or female?” Screenshot from YouTube.

Li Ting is in the business of entertainment. His exaggerated voice is amusing, and he makes his audience (in this case me and ARZ) laugh. What is not clear here is whether it was through Li Ting’s overly exaggerated, childlike, feminine vocal timbre or Li Ting’s specific pronunciation of certain words that the mainland Chinese gamer was able to deduce that he was Taiwanese. What is clear is that Li Ting is leaning into a cultural practice called *sajiao* (撒娇), or a “coquettish display of infantilized cuteness.” While *sajiao* does not usually have a comical effect, the *sajiao* Li Ting uses is over-exaggerated to the point that anthropologist DJ Hatfield (in response to my presentation of this topic the 2023 AAS-in-Asia conference in Daegu, South Korea) called it “sajiao drag.” Li Ting intentionally leans far into the *sajiao*, turning it into an exaggerated performance meant to draw attention not only to his voice, but to his listeners’ perception of his voice. As we watched together, ARZ explained why he laughed at Li Ting’s voice:

I just thought it was pretty funny, because when he started the accent just now, he opened his mouth and was clearly imitating how some Chinese people imagine Taiwanese pronunciation. Because I’ve seen a Taiwanese YouTuber, and the topic is what Chinese people think of Taiwanese pronunciation. So some people will purposely drag the voice out long, like the one just now. He’s very clearly imitating Chinese people imitating Taiwanese pronunciation, and I think it’s really funny.¹²⁶

¹²⁶ Interview with author, March 9 2023.

[我剛剛覺得還蠻好笑的，因為他剛剛一開始那個。開口明顯是在模仿某一些中國人想像的臺灣腔，因為我有看過有一個臺灣的 YouTuber，他在做的主題就是他去看那些中國人模仿的臺灣腔，然後有一些就是說故意把聲音拖很長，像他剛剛那個櫻桃就是。他很明顯是在模仿中國人想像的臺灣加，我覺得還蠻好笑的。]

While Li Ting's vocal acrobatics may be motivated by entertainment value, his voice here also becomes a commentary on certain preconceptions of Taiwanese Mandarin accents. He reclaims mainland Chinese perceptions of Taiwanese Mandarin accents by voicing an exaggerated version of these perceptions back to them. By over-identifying with an essentialization of the Taiwanese voice, he engages in a practice that Eidsheim, building on Gayatri Spivak's "strategic essentialism" and Jose Muñoz' "disidentifications," defines as "mimicry of racial mimicry" (Eidsheim 2019, 32; Spivak and Rooney 1989, 138; Muñoz 1999). Eidsheim suggests that focusing on this complex process can "open us to a type of micropolitics of listening, where the determination of race, essentialism, and naturalized concepts can be analyzed and contested by the listener as well as the vocalizer" (2019, 33).

By listening to Li Ting's voice, then, I can hear his funny vocal register and his efforts to create entertainment by engaging his mainland Chinese teammate in a discussion deeper than the typical videogame chatter. By listening to his mainland Chinese teammate's reaction to his voice, I can hear how stereotypes of Taiwanese accents play out in the context of videogames. By further listening to how Taiwanese and mainland Chinese listeners hear and react to Li Ting's voice, I can begin to understand how Li Ting's "sajiao drag," draws attention to certain perceptions and assumptions that Taiwanese and mainland Chinese people may have about each other, mimicking these perceptions to such an exaggerated degree that his listeners are forced to confront them head on.

These various levels of discourse exemplify the role of online videogames in cross-strait relations. Though hostility between Taiwanese and mainland Chinese gamers is perhaps the default, videogames can also provide a site to productively grapple with cultural, political, and linguistic differences across multiple layers of remediation. An exaggerated accent within the relaxed environment of online play adds entertainment value. It also provides a way for people to face their own biases, as well as the biases of their teammates and enemies. Videogames in this context are therefore much more than a site for play. They provide a site to try on a specific identity and to express one's beliefs in relation to others. With few opportunities for people in Taiwan and mainland China to interact, they provide a site for the complex negotiations of cross-strait relations to occur through the daily interactions of individuals on both sides of the Taiwan Strait.

5. Continuing the Cycle: Audio Communication and Regulation in the Age of AI

Throughout this dissertation I have discussed case studies of how people have used audio communication technologies to speak and listen across the Taiwan Strait despite restrictions. Until the 1980s, media in Taiwan and mainland China were tightly controlled, and yet communication persisted on a small scale through loudspeakers and radio broadcasts (Chapter 2). By the late 1990s, regulatory mechanisms in mainland China increasingly blocked shortwave radio signals just as the internet introduced new opportunities for unregulated interaction. Even as mainland China's Great Firewall developed to prevent communication with the outside world, platforms such as Clubhouse have briefly become sites for cross-strait exchange before the establishment of corresponding regulations (Chapter 3). The porous nature of mainland China's internet censorship also enables online communication on a small scale, as in the case of videogames where online gamers take advantage of a "gray zone" in regulation to access international versions of specific videogames and play together with other gamers outside of the Great Firewall (Chapter 4).

In this chapter I formulate these case studies into the framework of the repeating cycle of "regulatory latency" (also known as "regulatory lag") and explore how this cycle has continued as artificial intelligence (AI) technologies become more widely used in the context of cross-strait communication. I define regulatory latency as a cyclical pattern in which the development and implementation of regulating technologies often follow the widespread use of the communication technologies they aim to regulate.

In recent years, AI technologies have become increasingly central in both facilitating and regulating online communication. Debates surrounding the ethics of new AI technologies reveal that these tools often follow the cycle of regulatory latency, with the adoption of new AI technologies often preceding corresponding regulation.¹²⁷ While regulatory latency is frequently portrayed as a negative issue needing to be solved, here I position it as an oftentimes unavoidable process that is not inherently positive or negative. Instead, the process of regulatory latency provides the opportunity to explore what happens (both good and bad) in unregulated and less regulated communication environments.

Regulation of communication in the age of the internet looks very different on each side of the Taiwan Strait. Mainland China's internet is highly regulated, prioritizing the voice of the government at the expense of open communication and freedom of expression. The Great Firewall largely restricts access to the global internet, with mainland Chinese platforms dominating the information landscape to create an environment different from that of much of the rest of the world. In the context of mainland China's highly regulated media environment, the latency between the development of a new communication method and the implementation of corresponding regulation has often been intentionally or unintentionally exploited by people who are seeking opportunities for open communication.

In contrast, since the 1980s Taiwan has maintained a relatively free and open communication environment. Regulation in this context centers on combating disinformation,

¹²⁷ Legal scholar Nathalie Smuha acknowledges that there is already a "race to AI" and calls for a parallel "race to AI regulation" (2021). Policy researcher Araz Taeihagh discusses how regulators and governments have struggled to maintain regulation for rapidly developing technology due to informational gaps between technology companies and regulators (Taeihagh 2021, 144).

with Taiwan as a target for disinformation campaigns coming out of mainland China. With disinformation shown to lead to increased polarization of beliefs in online environments, detecting and regulating disinformation is essential to maintaining spaces for democratic discourse online.¹²⁸ These contrasting regulatory climates indicate that regulatory technologies and the regulation of technology are not in and of themselves value-loaded, but instead can serve as tools across various specific contexts. While the regulation of disinformation online is considered necessary by many in Taiwan to maintain a democratic media environment, in mainland China similar regulatory technologies are used to suppress opposition to the government.

Indeed, debates about regulation of technology have long reflected the presence of various competing interests. At the core of these debates is the extent to which new technologies should be adopted prior to the development of mechanisms to regulate them. Those on the entrepreneurial side will often argue that too much regulation can impede innovation.¹²⁹ On the opposite side, more cautious adopters of new technologies will advocate for regulatory mechanisms to be developed and implemented prior to or in tandem with the adoption of new technologies.¹³⁰ In the context of AI technologies, ethical concerns associated with lack of regulation often relate to “the technology’s potential to breach fundamental rights such as privacy and non-discrimination, to inadvertently nudge and manipulate people so as to hinder their self-determination, or to harm people’s safety and security” (Smuha 2021, 59).

The focus on regulation in this chapter sidesteps these debates, acknowledging that the development of new technologies necessitates a certain degree of freedom while at the same time requiring comprehensive regulation to mitigate risks. Instead, I bring attention to the results of regulatory latency, both good and bad, by exploring what happens during moments of unregulated communication. While lack of regulation could pose ethical or existential threats in certain contexts, throughout this dissertation I have highlighted how in the context of the Taiwan Strait and specifically mainland China’s highly regulated media environment, regulatory latency has in fact led to rare opportunities for open communication. This is in part because the mainland Chinese government has continually attempted to maintain a high degree of control over the information that its citizens can access, while the process of regulatory latency has in turn allowed for various gaps and openings through which people in mainland China continuously find ways to communicate with the outside world.

¹²⁸ See discussion of echo chambers in chapter 3, including Diaz Ruiz and Nilsson 2023.

¹²⁹ Economist Edythe Miller wrote in 1996 with the rise of cell phones and the internet, “A major rationale of the assault on regulation that has taken place in the work of orthodox economists over the past several decades, and in the courts and legislatures since the late 1970s, is the view that regulation retards technological advance and that an “optimal” rate of investment in new technology will be a happy by-product of deregulation” (E. S. Miller 1996, 719–20). This line of thinking continues today. As Allen and Berg explain, “There is now a growing consensus that the heart of a modern market economy is innovation” (Allen and Berg 2017, 218). Some entrepreneurs and scholars advocate for “permissionless innovation,” which “proposes that the default position for policymakers should be to allow experimentation and development with new technologies until demonstrable harms can be shown” (Allen and Berg 2017, 227; Thierer 2014). A concern is that if regulation is required at too early a stage in technological development, later stages of this development and implementation will be slowed by outdated regulatory requirements.

¹³⁰ For perspectives advocating for increased regulation of AI, see Yeung 2019 and Smuha 2021, and in the medical context Larson et al. 2021 and Minssen et al. 2020.

The idea of regulatory latency is often discussed in legal studies as regulatory lag.¹³¹ While “lag” and “latency” can be used interchangeably, here I choose latency specifically in reference to its usage in the audio context. In audio processing, latency refers to the delay between when an audio signal is input into a system (often via a microphone or instrument) and when it is heard through speakers or headphones. For online voice communication, latency can be defined as “the lag between speaking and hearing” (Williams et al. 2007a).

All online communication is limited by some degree of latency. In fact, information transmitted over the internet fundamentally cannot travel faster than the speed of light (Bozkurt et al. 2017). This means that even with the fastest internet connection speeds conceivably possible, online communication between someone in New York and someone in Paris, for example, would have a delay of 20 milliseconds due to the time it takes for light to travel. While online latencies of 20 milliseconds are generally accepted to be state of the art for activities such as online gaming and are imperceptible in the visual domain, in the context of music performance musicians have been shown to perceive latencies as low as two milliseconds (Jack et al. 2018). Perceptible latencies are then to some degree inherent in online audio communication.¹³² Latency in my understanding then implies a degree of unavoidable limitation.

Returning to the cycle of regulatory latency, while regulatory technologies and policies can at times preempt the implementation of corresponding communication technologies, these mechanisms are inherently limited prior to the development of the technology. Though it is possible to some extent to regulate a technology that does not yet exist, this process will always be limited by what remains unknown about the developing technology. In addition to delays in regulation, the concept of latency therefore also implies gaps, disruptions, or delays in communication. The idea of disrupted communication also extends metaphorically to the cross-strait context in which people in Taiwan and mainland China often encounter political and cultural differences while attempting to communicate online.¹³³

This chapter considers regulatory latency in the context of contemporary cross-strait communication. Having defined the concept, I first delve into the contrasting ethical stakes of voice surveillance, voice disguise, and detection of voice disguise in Taiwan and mainland China. Next, I focus specifically on the cycle of regulatory latency in the context of mainland China’s internet, identifying a pattern in which internet users continuously take advantage of gaps in regulation to find opportunities for open communication. Finally, I zoom out to consider the broader history of the development of AI for audio, exploring the stakes of this history for contemporary online audio communication and its regulation.

The ideas in this chapter are drawn in part from interviews with professionals who work on topics related to audio AI and/or disinformation. Of the 18 AI or disinformation professionals interviewed for this chapter, 13 were based in Taiwan, working either as professors or in industry.

¹³¹ For recent discussions of regulatory lag in the context of AI, see Ahern 2021; Taihagh 2021; Chhillar and Aguilera 2022.

¹³² These fundamental limitations of the speed through which information can travel online became a limiting factor for musicians hoping to perform together remotely during Covid-19 lockdowns. For a clear video explanation of this phenomenon, see: Neely, Adam. May 28, 2024. “What makes Meet the Grahams sound scary? Q+A.” <https://www.youtube.com/watch?v=X9RUh7w3i60>

¹³³ The concept of latency here relates to anthropologist Brian Larkin’s emphasis that media infrastructures do not always work smoothly and efficiently. The messiness, breakdowns, and discontinuity of media technologies can in turn lead to unique possibilities for interaction through unexpected uses of technologies (Larkin 2006).

To meet many of these researchers, I regularly traveled to the windy city of Hsinchu. Hsinchu is famous for its computer hardware manufacturing and home to Taiwan Semiconductor Manufacturing Corporation (TSMC), the world leader in the development and manufacturing of advanced semiconductor microchips. These chips are the backbone of computer hardware, providing the essential processing power for all types of digital applications, including AI. While the AI professionals I spoke with were not for the most part focused on hardware, their location at research universities in Hsinchu is representative of the centrality of hardware production in Taiwan's AI industry and professional community.¹³⁴ To supplement interviews with audio AI researchers, I also spoke with two people working on disinformation detection at audio-streaming platform Spotify. One of these people was employed to manually detect podcast disinformation in the local Taiwan context, and one was an engineer working on the automated aspect of this process, not specific to Taiwan. The initial phase of this research took place at UC Berkeley while I awaited a visa (delayed by COVID-19) to travel to Taiwan, during which time I took advantage of local expertise in Berkeley's Electrical Engineering and Computer Science (EECS) department. Conversations with professors there helped me to refine my research questions and provided a strong foundation for subsequent interviews and analyses conducted in Taiwan.

I complement the perspectives of AI professionals with those of internet users on the ground in mainland China and Taiwan. These perspectives were gathered while conducting research for other chapters of this dissertation, especially Chapter 3 (Clubhouse) and Chapter 4 (videogames). That these individuals are actively involved in cross-strait communication across various platforms means that they are also likely to interact in some way with the regulatory mechanisms of mainland China's Great Firewall. By integrating these diverse perspectives—ranging from end-users in Taiwan and mainland China to academic researchers and industry experts in Taiwan and the United States—this chapter considers how internet users' perceptions and experiences with audio communication technology and regulation in the age of AI aligns with or diverges from the perspectives of the people developing this technology.

Voice Disguise and Regulation Across the Taiwan Strait

Leading up to the 2024 presidential elections in Taiwan, the NGO Taiwan FactCheck Center identified various examples of "deepfake" disinformation campaigns attempting to interfere with the election. Deepfakes are an emerging technique for visual, image, or audio generation in which deep learning tools are used to produce content that realistically impersonates the voice or face of a real person. Deep learning is a subset of the AI field of machine learning, which involves training a computer on a large amount of data so that it can reproduce a desired outcome. Within the field of machine learning, neural networks are a foundational tool. The basis of a neural network is that it is fed an input (or, in some cases, many inputs), and it produces an output. Deep learning involves training complex neural networks with multiple layers (called deep neural networks) to automatically detect features in large amounts of data. For a deepfake, the input might be audio recordings or video footage of a real person, and the output would be new content generated with that person's face or voice. The use of deepfake

¹³⁴ TSMC held over 60 percent of the global market share in semiconductor manufacturing as of May 2024, supplying its advanced microchips to advanced AI developers such as Nvidia (J. Lin 2024).

technologies to produce content for purposes such as celebrity impersonations and entertainment has raised ethical questions about attribution and copyright. Deepfake technologies have also been used to spread disinformation, which increasingly poses a threat in global politics.

Taiwan's 2024 presidential election took place on January 13. Incumbent president Tsai Ing-wen of the Democratic Progressive Party (DPP) was ineligible for reelection due to term limits, so the DPP nominated then Vice President Lai Ching-te, who went on to win the election. He ran against the opposition candidates Hou Yu-ih of the Kuomintang (KMT) party and Ko Wen-je of the emerging third-party Taiwan People's Party (TPP).

In August 2023, a fake audio recording circulated in which Ko Wen-je purportedly criticized Lai Ching-te's visit to the United States. The content of this audio file was quickly debunked by Ko's own party and Taiwan's Investigative Bureau. The Taiwan FactCheck Center found that this content was most likely synthesized using deepfake technology. In November 2023, a fake video circulated showing Lai Ching-te allegedly expressing support for an alliance between the other two parties' candidates. The Taiwan FactCheck Center found this video was created using both the manual manipulation of images (known as "cheapfakes") and voice generation with deepfake technologies.¹³⁵ Since the lip movements in the video did not match up with the spoken words, the video was most likely created by reusing old video footage paired with a deepfake audio track. In another recording circulated one month prior to the election, a deepfake voice impersonating Lai Ching-te made improbable negative comments about Lai's own party (W. Li 2023).

While this type of media may appear as obviously false for a Taiwanese audience, the Taiwan FactCheck Center suggested that it may be more difficult to detect flaws in foreign content. One such video circulated in Taiwan in late December 2023, and consisted of US House Representative Rob Wittman supposedly endorsing Lai and his running mate Hsiao Bi-khim. The Taiwan FactCheck Center reported that the video was "most likely uploaded by a TikTok private account before being reposted on Reddit, Facebook and PTT, a popular Taiwanese online forum." The Facebook post connected to this video stated, "This is the first time that the United States has spoken out to choose a side" in the Taiwanese election (W. Li 2024). Taiwan FactCheck Center found that the footage in the video was adapted from a television interview by Whittman in which he talked about the war in Ukraine, and then deepfake audio content was dubbed over the video. Since a Taiwanese audience is not familiar with US politics or Whittman, it was difficult for many "to discern the flaws in the fake video" (W. Li 2024). The Taiwan FactCheck Center further traced the circulation and comments on the video and found these were likely produced through coordinated disinformation efforts. Much of this content originated in Facebook fan pages operated outside of Taiwan, and accounts of the people commenting on the videos were created within a short time span and had limited account activity (W. Li 2024).

¹³⁵ Paris and Donovan distinguish between "deepfakes" and "cheapfakes." They define cheapfakes as "manipulated images created with more conventional techniques such as photoshopping" and deepfakes as "visual or audio files that were produced through machine learning processes to generate images of events that look real" (Paris and Donovan 2019; W. Li 2023).



A screenshot of the fabricated video in which the DPP presidential candidate Lai Ching-te appeared to support the cooperation of the KMT and the TPP party in the 2024 Taiwanese presidential campaign. The Taiwan FactCheck Center has debunked this video as untrue.

Figure 17. Screenshot and caption courtesy of the Taiwan FactCheck Center. Reproduced from <https://tfc-taiwan.org.tw/articles/10025>.



A screenshot of a Facebook post shared a TikTok video titled "American online video on the 29th: Vice Chairman Rob Wittman on the House Armed Services Committee was interviewed and openly endorsed the Democratic Progressive Party." The short video was based on a clip from a U.S. local news station, WUSA 9, dubbed with AI synthetic voices to make a statement that the congressman never said.

Figure 18. Screenshot and caption courtesy of the Taiwan FactCheck Center. Reproduced from <https://tfc-taiwan.org.tw/articles/10025>.

Deepfakes are simply the latest strategy in foreign disinformation campaigns in Taiwan, complementing efforts largely initiated by the mainland Chinese government to mislead the Taiwanese public, sow distrust in government, and heighten polarization (Doublethink Lab 2024; Hsu 2024). Since these campaigns do not depend on a high degree of accuracy in terms of the disinformation content, it is not immediately clear the extent to which deepfakes will further the goals of disinformation campaigns in the long term. As Andrew McDowell, a disinformation researcher at Spotify, explained, people are often influenced by information that seems to be obviously false, so having more realistic representations of fake information will not make much of a difference. McDowell said:

In terms of disinformation, I don't feel it would dramatically make things worse than they are now. If you put out a perfect video tomorrow of Biden saying something absolutely awful, the same people who are going to believe it would believe you putting out a badly edited clip. I don't think it fundamentally changes things. I think there are places where it could be very dangerous. But I think on a largescale people already trust text way too much...and the same people who are going to fall for deepfake video are going to fall for it in text.¹³⁶

Indeed, much of the disinformation circulated in Taiwan is obviously falsified, as in the example where a deepfake of Lai's voice spread negative content about his own political party.¹³⁷ The Taiwan FactCheck Center also found that most audiovisual disinformation circulated prior to the 2024 election were still “based on existing old videos or audio files repurposed with simple techniques” such as photoshopped images or images taken out of context to support a false or misleading claim. Even so, Taiwan FactCheck Center clarified that even if AI generated disinformation remains less common than less advanced techniques, “this does not mean we can shrug off experts' warnings about the challenges imposed by deepfake and AI technology” (W. Li 2023). This is because the risks of rapidly advancing disinformation techniques are as yet unknown in the context of Taiwan's media environment.

Deepfakes are a new and developing technology, and comprehensive detection and regulation have yet to be developed and implemented. According to UC Berkeley computer science professor Hany Farid, there are a number of easily accessible online tools through which anyone can produce an audio deepfake. Deepfake voice detection, on the other hand, requires a high skill level with limited availability of reliable public detection tools.¹³⁸ In other words,

¹³⁶ Interview with author, August 10 2023.

¹³⁷ In another example, one campaign identified by the Doublethink Lab featured videos where a news announcer's mouth was fully covered by a large round microphone, making it impossible to verify whether the announcer was actually pronouncing the words in the video (Author's interview with WCP, November 9 2022).

¹³⁸ In an interview with reporter Lauren Leffer for *Scientific American*, Farid said, “There are services that you can pay \$5 per month for [that let you] upload your reference audio and clone the voice. Then you can type and get convincing audio in a few seconds. This is text-to-speech. There's also a second way to do this called speech-to-speech. I record a person and clone their voice. And then I record myself saying what I want them to say with all the intonation—bad words and all—and it converts my voice into their voice. It's all the same underlying generative AI technology. For either method, anybody can do this. There is no barrier to entry or technical skill involved... There's a huge asymmetry here—in part because there's a lot of money to be made by creating fake stuff, but there's not a lot of money to be made in detecting it. Detection is also harder because it's subtle; it's complicated; the bar is

regulatory technologies for deepfake voice lag behind technologies that produce deepfake. Deepfake technologies then fit into the ongoing cycle of regulatory latency as a new communication technology where corresponding regulation has yet to be implemented.

The continuity of regulatory latency becomes more evident within the broader trajectory of voice identification and voice disguise technologies. In the context of mainland China, state level voice surveillance and identification has become increasingly common. Journalist Mara Hvistendahl reported in 2020 that large technology companies, especially a voice-focused one called iFlytek, are at the forefront of this development. iFlytek produces software tools for everyday use, such as voice dictation and live voice translation for travelers. These useful (and widely used) tools then enable the company to collect large amounts of data about its users. iFlytek also collaborates with the mainland Chinese government and cooperates with the agency that operates prisons in Xinjiang, a region heavily populated by the predominantly Muslim Uighur minority group that has faced persecution and widespread detainment in internment camps by the mainland Chinese government. Hvistendahl further revealed that iFlytek collects language data on regional and less common languages and dialects, ostensibly to develop technologies inclusive of these languages. However, their partnership with the government and their focus on Tibetan and Uighur languages—two groups that have faced systematic oppression in mainland China—suggests possible ulterior motives related to surveillance (Hvistendahl 2020). Human Rights Watch has reported that iFlytek’s voice biometric technologies and collaboration with mainland China’s Ministry of Public Security pose risks for individuals, including surveillance and retaliation against those critical of the government (Human Rights Watch 2017).

The use of voice identification for surveillance is not limited to mainland China. Media theorists Jonathan Sterne and Mehak Sawhney write, “Drawing on the recent boom in artificial intelligence and more specifically machine learning, a growing machine-listening industry has arisen that is predicated on mass surveillance and the expropriation of data from people without their knowledge or full consent” (Sterne and Sawhney 2022, 289). This “machine listening industry” has led to the increased prevalence of voice identification as a biometric technique used by both governments and technology companies. For example, in 2018 Amazon filed a patent to analyze and recognize users’ accents to determine their ethnic origin (Pfeifer 2023; B. Lin 2018). Similar technology has been used to identify where people come from in asylum cases, and has led to incorrect conclusions; for example, a case in Germany caused someone to be incorrectly denied asylum based on his accent (Pfeifer 2023; Biselli 2020). In US prisons, voice prints of incarcerated people are collected and extracted to build databases of calls that include the voices of people on the other end of the line (Pfeifer 2023; Joseph and Nathan 2019). Media scholar Michelle Pfeifer writes that errors can have huge consequences in these settings, exacerbated by the fact that “algorithmic or so called intelligent bordering practices assume neutrality and objectivity and thereby conceal forms of discrimination embedded in technologies” (Pfeifer 2023).

Voice identification efforts can be subverted through voice disguise. A mainland Chinese internet user concerned about surveillance might use an online filter to change the pitch of their voice, for instance to make it sound high and nasally. Radio Free Asia reported in 2021 that the

always moving higher. I can count on one hand the number of labs in the world that can do this in a reliable way. That’s disconcerting” (Leffer 2024).

use of voice disguise by mainland Chinese internet users on the audio social media app Clubhouse (discussed in Chapter 3) led the Chinese Cyberspace Administration to call on tech companies to increase efforts to counteract these strategies (Long et al. 2021). With simple voice disguise techniques that involve changing only one or two spectral or temporal characteristics, it is increasingly possible to reverse the disguise and identify the voice. When voice disguise is more complex, as with voice conversion techniques where multiple parameters are changed in order to imitate another person's voice, it can be more difficult to detect the disguise (L. Zheng et al. 2021; M. K. Singh 2024).

Deepfakes differ from other voice disguise techniques because they are generated from scratch (rather than altered from a recorded voice) and use deep learning to imitate a real person's voice. While voice detection tools identify people based on their voices, with the spread of deepfakes it also becomes necessary to determine the presence or absence of deepfakes. This is because with the spread of deepfake technology it is now reasonable to assume that statements made by a specific voice, even the voice of a real person, may have been falsely generated. Deepfake detection is then necessary to distinguish real from fake and true from false. That said, deepfake audio is as yet difficult to detect and regulate. In a 2023 study, Farid's group at Berkeley reviewed and proposed various deepfake detection methods while also emphasizing the need for companies developing deepfake technologies "to help mitigate potential abuse from deepfakes by embedding imperceptible watermarks into synthetically generated content." (Barrington et al. 2023).¹³⁹ This type of initiative is unlikely to be widely adopted without adequate policies, advocacy, or incentives. As yet, there are limited policies enforcing deepfake detection requirements, meaning that it is not only the development of regulating technologies that lag behind, but also the implementation of policies at a governmental level.¹⁴⁰

While the lack of deepfake detection and regulation pose threats for the spread of disinformation in Taiwan and beyond, I have shown through the framework of regulatory latency that new and as yet unregulated technologies can often provide opportunities for open exchange in highly regulated media environments. In mainland China, voice identification and detection of non-deepfake voice disguise are becoming more comprehensive, even while deepfake detection lags behind. This means there is a potential for individuals to use deepfake to disguise their own voices as a way to avoid being identified.

For this reason, Taiwanese speech and AI researcher ZT said that when discussing sensitive information, he would choose to use voice rather than text. He explained,

If I want to talk about some sensitive topics or talk to good friends when my students are not supposed to hear, usually we don't use text. I'll say let's call or let's talk...I think it's more secure. And now we have an excuse because we have deepfake. If there is a recording saying something, I can say 'Okay it's a deepfake, it's not me' [laughs]. 'It's not me, it's someone else.' Because we have the voice conversion

¹³⁹ They suggest Adobe's Content Authenticity Initiative as an example (Barrington et al. 2023).

¹⁴⁰ In the context of Taiwan, Taiwan FactCheck Center reported that the government is working on a draft of basic AI law. Wei-ping Li writes that current policies work to "punish actors who intentionally spread false information to influence the election results. However, in practice it would be hard to locate the creators and spreaders of the false information, not to mention produce evidence of the actors' intentions. As AI has shown its potential to disrupt elections, politics, and society, it would be imperative that Taiwanese society develop more tools to tackle the problem" (W. Li 2024).

technique, so I know I know what we can do. So I can use [deepfake] as an excuse [to avoid blame or detection].¹⁴¹

Interestingly, ZT's perspective is not that voice is harder to identify than text. Instead, he claims that deepfakes are difficult to distinguish from real voice, blurring the line between true and false and providing a way for individuals to avoid blame and conclusive evidence over something that was said. The lack of comprehensive deepfake detection and regulation poses a significant risk in many contexts where it is imperative to identify true versus false information. At the same time, in this current moment where deepfake communication technologies precede the development of comprehensive detection tools, deepfakes can provide cover for people in restricted communication environments searching for the opportunity to communicate openly without being identified.

Leveraging Regulatory Latency for Open Communication in Mainland China

Internet users in mainland China have long found ways to avoid detection by leveraging gaps in the regulation of communication technologies. One current strategy, for example, involves communicating with less common languages for which data is not readily available. Another approach is to use parody and humor, which automated AI detection tools notoriously struggle to interpret. These strategies constitute examples of “obfuscation,” defined as the “deliberate use of ambiguous, confusing, or misleading information to interfere with surveillance and data collection” (Nissenbaum and Brunton 2015, 1). Obfuscation tactics can be employed to protect privacy, avoid surveillance, or prevent unauthorized access to personal information.

Obfuscation techniques often work in tandem with cycles of regulatory latency. In the context of mainland China's internet, data scientist Jason Q. Ng writes, “Over the years, in a series of cat-and-mouse games, Chinese internet users have developed an extensive series of puns—both visual and homophonous—slang, acronyms, memes, and images to skirt restrictions and censors” (2013). Thus far I have discussed regulatory latency in the context of various communication mediums: loudspeakers, radio, social media, and videogames. However, the development of obfuscation techniques on mainland China's internet suggests a separate cycle of regulatory latency. As Marshall McLuhan writes, “The content of any medium is always another medium” (1994).¹⁴² The content of the internet is thus composed of text, images, and speech—all of which are mediums in and of themselves, and all of which require unique tools to be regulated.

A classic obfuscation method in the realm of text is to use homophones, or Chinese characters with similar sounds but different meanings. Homophones were widely used to avoid some of the earliest forms of text censorship by replacing banned characters with alternatives that were pronounced similarly but written differently. This meant an informed reader would understand the hidden meaning, but the content would be less susceptible to censorship. Hvistendahl explains:

¹⁴¹ Interview with author, November 15 2022.

¹⁴² This concept has been termed “remediation,” famously discussed by Bolter and Grusin 2016. See discussion in Sterne 2012, 9.

In the early 2000s, as online censors banned certain characters, computer users got around the state by switching to homophones. To mock the notion of a “harmonious society,” a Maospeak phrase popularized during Hu Jintao’s rule, they joked about crustaceans – *hexie*, river crab, is pronounced similarly to *hexie*, harmony. “Serve the people” became “smog the people”...Alarmed by the proliferation of online sarcasm, the central government went so far as to ban homophones and other wordplay. So dissidents turned to other means of dissemination (2020).

As the censorship apparatus caught on to this technique in the early 2000s, internet users cycled on to the next medium, embedding text within image files and sometimes turning images on their sides. Easily interpretable to a human eye, these strategies also easily fooled early automated detection mechanisms that were less advanced at image processing than text processing.¹⁴³

When image processing improved, internet users switched to video and audio communication to avoid automated detection. However, even these methods are increasingly subject to surveillance. Hvistendahl writes, “As the CCP has amped up its surveillance capabilities over the past decade, it has installed millions of cameras, introduced electronic ID cards and real-name registration online, and built tech-driven “smart” cities” (2020). People hoping to avoid detection can use strategies such as covering their faces, but with face ID frequently required it is increasingly difficult to avoid surveillance in daily life.¹⁴⁴

Widespread online video and audio surveillance in mainland China means that internet users have become even more strategic in their obfuscation strategies. Current strategies of using uncommon languages and humor to avoid automated detection mechanisms demonstrate an intuitive understanding of how the latest communication and surveillance technologies work and how the shortcomings of these technologies can be leveraged for open communication.

Uncommon Languages as an Obfuscation Technique

When mainland Chinese internet users speak with uncommon language as an obfuscation strategy, they are leveraging specific latencies in the data collection process. The first step for many automated speech technologies (including speech censorship and detection of audio disinformation) is to transcribe the speech (convert speech to text). Developing speech to text technologies requires a huge amount of data for each specific language and variation of that language. Since there is more available data for common languages such as Mandarin and English, speech to text tools often work better for these widely spoken languages. At the same time, a lack of comprehensive data for less common languages, dialects, and accents means that these linguistic variations may be less subject to censorship.

Chinese censorship researcher BY (who currently lives in the United States) said that he will sometimes change his accent or dialect in order to avoid censorship when using the Chinese

¹⁴³ Jason Q. Ng writes, “The censors are not infallible, and it is possible for posts with banned words to escape the censor’s eye—so long as they don’t gain too much attention or advocate collective action, or perhaps if they’re cleverly embedded inside images or obscured in coded language” (Ng 2013).

¹⁴⁴ While there have been measures taken by the mainland Chinese government to curb the overuse of facial recognition technology, this form of biometric identification remains widespread (Ye 2023).

social media site WeChat, since in his experience speech to text capabilities are less comprehensive for languages other than standard Mandarin:

I'm not quite sure how powerful the censorship engines are. I know they have pretty good voice recognition AI for standard Mandarin. If you're from Beijing you probably speak standard Mandarin, but I'm not from Beijing, so sometimes when I speak about sensitive issues with my friends, I just use my hometown dialect in order for it not to be recognized. Because if you have a WeChat account and you send a voice message, you can convert the message into text. Some people don't want to listen to those voice messages, so they just convert them into text messages most of the time. But when you're using your own accent, using your local language, [WeChat] cannot [convert text to speech] very well. But I know that they are working on that because I know not only WeChat, not only Tencent, but also some local governments are hiring people who can speak different local languages. They're working on that. So it's a matter of datasets and a matter of time.¹⁴⁵

From his experience as a WeChat user, BY intuits that speech to text and censorship capabilities for his hometown dialect are less advanced than those for standard Mandarin. However, since he follows the industry closely, he also realizes that this situation is rapidly shifting in the continuous cycle of regulatory latency, with tech companies increasingly prioritizing expertise and data collection of local languages.¹⁴⁶

The obfuscation strategy of using uncommon languages and dialects to take advantage of gaps in data collection may work especially well for speech, and especially well in the Chinese language context. This is because both speech communication and Chinese languages are diverse and varied. Andrew McDowell explained the challenges of data collection for speech in the context of Ireland's languages (where he is from):

Languages and accents, things like that, are much more of a challenge in audio than they are in text. Everyone in Ireland fundamentally writes the same way. But if you get someone from Cork and someone from Dublin, they sound like they're speaking different languages. The accent is remarkable. It's incredible. And that's a real challenge. Pronunciation varies so much from region to region.

This dynamic holds true for the Chinese language context as well. Chinese language variations often have identical written forms but different pronunciations. For instance, the Mandarin word for "eat" (吃) is pronounced "chi," while in Taiwanese Hokkien it is pronounced "jia." While in writing "eat" is just one character, in spoken language there are countless variations of this one word. These variations mean that more data needs be collected to develop comprehensive

¹⁴⁵ Interview with author, October 7 2021.

¹⁴⁶ It is not only Chinese companies that are prioritizing collecting data on less common languages, but also US companies. As Sterne and Sawhney discuss, projects such as Google's "Next Billion Users" aims to bring people from developing countries online in order to expand its user base, while also collecting data about people without proper consent (Sterne and Sawhney 2022, 296–98).

automated speech detection tools, since an AI engineer would need to train a model on data that encompasses all of these variations.

There is a long precedent in mainland China, going back to the 1950s, of promoting linguistic assimilation under the guise of national unity (Sato 2022). The context of automated speech censorship reveals another motivation for promoting Mandarin as a national language at the expense of the suppression of linguistic diversity: if everyone speaks the same language, comprehensive surveillance is easier to establish, since there is no need to collect diverse speech data for varied accents and languages. At the same time, the widespread use of diverse dialects and languages in China provides an obfuscation technique to get around surveillance, as people take advantage of the longer latencies in gathering diverse audio data by using these languages to communicate outside of the bounds of the latest regulatory technologies.

Even while speech applications of AI require more varied data, AI professionals often comment that the process of data collection is slower and more challenging for audio than it is for text. One professor in Taiwan said that though she has a personal interest in music and audio, she has never pursued AI audio research because the data collection process seemed too time-consuming and labor-intensive. Her current research involves text data scraped from Twitter. If she were to focus on audio data from YouTube videos instead, she said that processing a single video would take a few minutes. However, in the same amount of time she can scrape “tens or thousands of the tags (on Twitter), so the quantity definitely will not be the same.”

Similarly, McDowell elaborated that the process of building and labeling one’s own datasets, as is often necessary for academic labs or initial research, is more time consuming for audio than for text. Once the dataset is built, McDowell said that it often takes more resources to process audio datasets than text datasets. For this reason, he suggested that in many cases it makes sense to develop new deep learning models and take experimental risks with text first, and then later apply these models to audio data:

Why text comes first is just because it’s easier...It’s just harder to build your own datasets in audio. If I’m doing an experiment, I can take 300 text examples, and if they’re relatively short I can label them myself in a few hours. I can’t do that with audio. It’s just that you have to listen to it. And that’s the killer with all of this. You can’t browse, you can’t scan. You can listen to things at high speed, but if you want to have a human label audio, they have to listen to it. It just takes longer. So it’s a challenge to build your own datasets with audio and label them.

You can process text extremely cheaply in huge volumes on not very expensive hardware, but doing audio does require more of a commitment in terms of hardware and time... it’s much harder to commit that kind of experimentation to audio when you could build a huge data set, and then it doesn’t work. With text, it just doesn’t take as long to build your data set. You can afford to take risks and a sort of a faster pace there.¹⁴⁷

The latencies associated with processing audio data are also perceived by internet users on the ground. Taiwanese reporter Liya Chen, who spent three years living and reporting in mainland China, said that she noticed that for audio it often takes longer for messages to be censored on Chinese social media sites. She explained:

¹⁴⁷ Interview with author, August 10 2023.

I think if there's sound on TV, it takes more time, and maybe needs a little more computer memory, so on a technical level sound will be slower than text. But that doesn't mean that the censorship standards are different between the two. Because in reality, what I've heard and experienced is if you call someone in China and you are talking about a slightly sensitive topic, you will realize there are some strange noises showing up, or it will suddenly disconnect. This means that their machine is still censoring, still operating, just maybe in comparison it's slower than text. For example, if today I sent a message on WeChat saying 'Taiwan independence,' the receiver might not see it, because as soon as I send it, it has already been placed as sensitive vocabulary. But if today I use voice to say 'I think Taiwan should be independent,' they are still censoring, but I think there's a time lag.¹⁴⁸

我覺得聲音在電視的時候，他需要比較多的時間，可能記憶體也要大一點，所以在技術層面上來講，聲音會比文字慢。可是並不是說兩邊的審查的標準有沒有什麼差別。因為也的確像是說，也的確上有遇過，你打電話給對方在中國那邊的人，你們在聊一個有一點敏感的話題，但你會感覺到已經出現了一些奇怪的雜音，或者是突然被斷線。那其實這個就表示說他們那個機器在審查，還有在運作的，只是可能相對來講比較慢，就是它比文字更慢。比如說我可能今天打一個【臺灣要獨立】在微信上有可能對方根本就看不到，因為我一發出去的過程，他就已經進到了所謂的敏感詞彙。可是如果我今天是用聲音講說，我覺得臺灣要獨立，他還在審查，所以我覺得會有一個時間差。

While speech is certainly subject to censorship in mainland China, Liya's perspective demonstrates the perception that speech censorship can happen more slowly than text censorship. These perceptions align with AI professionals' experiences of the comparatively longer time it takes to build and process audio datasets.

Perhaps because gathering and labeling audio data is more time consuming, publicly available text datasets have existed for longer and are often more accessible than audio datasets. In addition to the amount of time it takes to collect and process audio data, McDowell explained that it can be difficult to even access audio content online:

There's certainly a lot more text examples out there that are just publicly available...Scraping text off the internet is easy. OpenAI is just scraping the internet, that's how they're building their models. They've done that for audio as well. But it's harder to do...Unless you're a platform yourself—unless you are YouTube who is uploading all the videos—it's not so easy for someone else to go and get that audio and say, 'This is a collection of the conversations going on in Taiwan today.' So there's the mechanics of it, that text data is much easier to acquire. For text, it's just websites. You can check IP addresses. You can check domains, where things are registered. Audio tends to be on a platform which may expose that information, or they might not. On some platforms it's very hard. A lot of them are just closed box completely.¹⁴⁹

Taiwanese disinformation detection civil society organization Doublethink Lab has run into this type of challenge while tracking disinformation on YouTube. For text and image

¹⁴⁸ Interview with author, May 20 2022.

¹⁴⁹ Interview with author, August 10 2023.

disinformation on Facebook, they were able to sign a contract with Facebook to monitor the data. However, this process proved to be more difficult for audio-visual content on YouTube. Doublethink Lab founder Puma Shen remarked,

We actually signed a contract with Facebook so we can look into the data from the back end. So that's why we can really look into the archive data and see what happens. Lots of them are just tags or images so it's easier to monitor Facebook. But on YouTube it's impossible. You need lots of disc space if you want to download those videos. You need a bigger company to do that kind of stuff.¹⁵⁰

McDowell clarified, however, that it may be just a matter of time until audio datasets are widely and publicly available, even for independent researchers and small organizations. In recent years, language detection tools have been developed to make tasks such as converting speech to text more efficient even for people outside of large tech companies, who have fewer resources in terms of time and computer power. In the context of some of the US's largest tech companies, OpenAI's Whisper was released as open source in September 2022 as an automatic speech recognition (ASR) system designed to transcribe and translate speech in multiple languages. It was developed with training data that OpenAI scraped from across the internet.¹⁵¹ Wav2Vec, released as open source by Meta AI in October 2019, was also groundbreaking in automatic speech recognition for its use of convolutional neural networks (CNNs) and transformer architecture to process raw audio (wav) data, representing progress in the efficiency of audio processing and significantly reducing the amount of labeled data required to develop automated speech recognition tools.¹⁵² These tools will likely make it easier for independent researchers and small organizations to build speech tools for purposes such as disinformation detection. They may also lead to more comprehensive censorship tools even for less common speech variants, meaning that the use of uncommon languages for obfuscation may soon become less effective.

Humor as an Obfuscation Technique

The use of humor, irony, sarcasm, and parody can be considered a classic obfuscation technique on the Chinese internet, used across the mediums of text, images, and speech. Sociologist Guobin Yang suggests that humor and satire constitute a part of the emotions that are central to online activism in China, with these strategies especially significant in questioning centralized power and reverence for communist party officials:

It is against this culture of official-centricity that the internet culture of humor and play assumes special significance. Play has a spirit of irreverence. It always sits

¹⁵⁰ Interview with author, December 15 2022.

¹⁵¹ McDowell described OpenAI's Whisper: "Language detection has gotten a lot better in audio in recent years. Like the Whisper model which OpenAI released. It's a speech to text model, but it can detect language with remarkable accuracy. It's phenomenal. Whisper is just an absolute game changer for anyone trying to do this themselves. And it's remarkably effective" (Interview with author, August 10 2023).

¹⁵² Interview with author, August 10 2023.

uncomfortably with power... Much online activism, and much Chinese internet culture in general, is enlivened with this spirit... In Chinese cyberspace, nothing is sacred. Pretensions to authority are favorite targets of attack (G. Yang 2009, 224).¹⁵³

For humor in text specifically, Ng wrote in 2013 that automated detection tools were not equipped to detect ironic comments on the Chinese social media site Weibo. Instead, these comments were left to human censors employed by Weibo to manually sift through and detect. Ng explained:

Chinese internet users have mastered the use of irony as protest, reaching the point where emphatically pro-government comments online such as ‘socialism is good’ and ‘I have been represented by my local official’ are often meant to be satirical. Filtering tools including the [automated] ones Weibo uses in its search engine certainly can’t recognize such subtleties. In some respects, the filters are ‘easy’ to defeat, emphasizing just how important those human monitors employed by Weibo are (Ng 2013, xxi).

Humorous images have also been used as a vehicle to express social and political critique in mainland China. Journalist An Xiao Mina discusses what she calls “social change memes,” in which Chinese activists take advantage of the “best qualities of internet memes – rapid virality, irreverent humor, and participatory culture,” to embed activist messages within seemingly innocuous media. Mina writes that these memes are difficult to censor not only for automated censorship tools, but also for human censors: “Even to a trained human censor, a picture of a llama or a crab could be subversive, or it could be a photo from the zoo. Deleting too many of the latter would cost political capital” (Mina 2014, 364).

Given that speech already poses challenges for online content moderation (due to speech variations and time-consuming data processing, as discussed above), voice communication combined with the obfuscation strategy of humor may be especially effective for avoiding censorship. JN, a Chinese PhD student studying in the US, described the use of satire in voice communication on the audio social media app Clubhouse (see Chapter 3). He said that censorship is often quite comprehensive and advanced, to the point that satire remains one of the few strategies through which people can express critique:

Part of it was just internalized censorship. On the internet, the only way that you can express some kind of dissent is by satirically praising. We see that as well in the COVID policy protests, where there’s so much human moderation. The amount of human manpower they put into online moderation—it’s unthinkable. Even any kind of code language will quickly become decoded and targeted by censors. The only thing they can do is just to write “hao” [good] under every piece of news about the Covid lockdown. You see a bunch of people just commenting, “hao.” And that’s obviously satirical. But it’s also serving a more subversive purpose. Due to the nature of censorship that’s been going on, especially the intensity of it in recent years, mounting a serious political

¹⁵³ See also (G. Yang 2018, 1946).

critique is very rare because it would almost instantly be censored. So people have to find increasingly innovative and subversive ways.¹⁵⁴

The “internalized censorship” that JN describes indicates that mainland Chinese internet users are so accustomed to disguising their political critiques that this type of coded language has become second nature.

While this type of communication may be intuitive to mainland Chinese internet users, outsiders often struggle to understand the nuances. In the Clubhouse chatrooms that JN joined, participants took turns praising political figures. The praise was so over the top that anyone familiar with mainland Chinese internet culture would quickly realize that the praise was satirical. However, JN said that sometimes people from Hong Kong or Taiwan would enter these satirical rooms. When they first joined, they might not understand the satire, so they sometimes spoke up and tried to disagree with the mainland Chinese Clubhouse users. JN recalled:

When you first join, the room title literally says, “Fan Club of Tu Zhuxi,” and you go in and you listen to all these people giving this extravagant praise. Now, I think for most people just listening to that for more than five minutes it would become very clear that it’s all just satirical roasting. But some people first join and they don’t understand, and they go up and they say they don’t like it. They don’t understand, so they think it’s all sincere. And then the host just says, “This is a kind of performative art for us. So we just ask you to please stay for another ten minutes...Please don’t take it seriously. Listen some more and you’ll see.”¹⁵⁵

Non-Chinese internet users are not the only ones who struggle to interpret this type of internalized “satirical roasting;” this type of communication also poses challenges for automated AI detection mechanisms. As JN said, “I don’t think AI is complex enough to be able to detect three layers of sarcasm.”

Indeed, current AI tools often struggle to understand even basic humor, to the amusement of internet users worldwide. For example, in May 2024 Google released a new AI-powered search feature called “AI Overviews” that summarizes search results. This new feature botched interpretations of satirical content, conveying to readers summaries of search results that were obviously false. In one example, AI Overviews was asked the satirical question, “How many rocks should people eat?” It responded, “Eating the right rocks can be good for you because they contain minerals that are important for your body’s health,” even going so far as to attribute this answer to UC Berkeley geologists. Google’s explanation for this clearly false answer was that there is not much web content on the topic of eating rocks, but there is satirical content. In this case, a satirical article from the joke publication *The Onion* with the headline “Geologists Recommend Eating At Least One Small Rock Per Day” was republished on a geology website, and the AI Overviews feature incorrectly interpreted the information in the article to be true rather than satirical. A similar example occurred when someone asked AI Overviews how to get

¹⁵⁴ Interview with author, January 30 2023

¹⁵⁵ Here JN mentions Tu Zhuxi as one example of possible subjects for these types of parody rooms. Tuzhuxi (兔主席) is a famous online influencer in mainland China who is supportive of the CCP government. See Chapter 3 for a discussion of his presence and influence on Clubhouse. Interview with author, January 30 2023.

cheese to stick to your and AI Overviews suggested mixing glue into the sauce, pulling this answer from a satirical post on Reddit from 11 years prior (Ladden-Hall 2024).¹⁵⁶

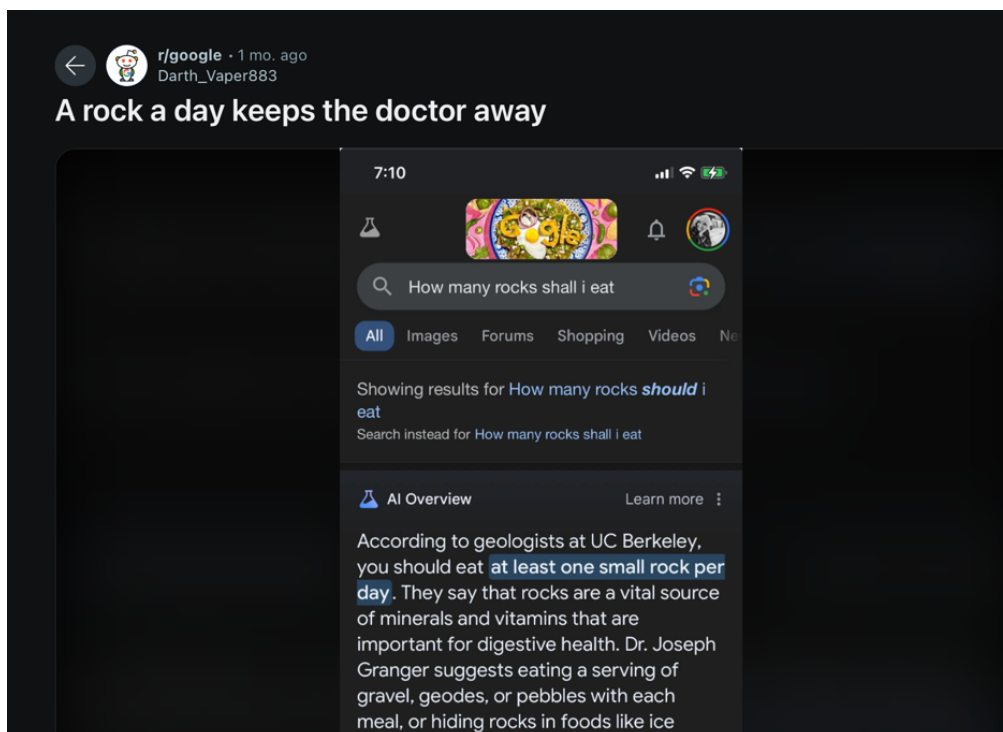


Figure 19. Screenshot of a Google AI Overviews mix-up, reposted to Reddit.

One issue behind these AI mishaps is that current AI tools struggle to detect context and tone. This difficulty stems from the fact that current AI language models are most often trained on the semantic meanings of words and the structural rules of language; they are trained on the literal, textual meaning of words rather than the nuances of spoken language. UC Berkeley Engineer Gopala Krishna Anumanchipalli explained, “The language that we speak is heavily structured, and right now there’s a monologue. I’m generating this stream of words that I’m formulating in my head, and it’s largely constrained by the question that was asked... So I can only say things about the development of AI for speech in this conversation, and not something about pizza.”¹⁵⁷ However, the meaning of speech is also influenced by factors such as context, intonation, speed, and facial expression—information that can be lost with statistical models trained on the literal meanings of words.

Contextual clues and tone of voice are especially pertinent to spoken language. As JN explained, a lack of contextual understanding can lead someone to misinterpret satirical praise as

¹⁵⁶ Was the AI overviews feature in on the joke, responding to an absurd question with an absurd answer? Perhaps. That said, the tone of the response was practical and serious, lacking the satirical subtleties that humans usually use to detect humor.

¹⁵⁷ Interview with author, November 15 2021.

sincere. However, information about tone of voice is often lost in AI models that convert spoken language to text. Anumanchipalli elaborated:

The same chunk of speech may make sense in one context, and in another context it could mean something that's totally different. And these are things that you cannot codify because it relates to, what is intelligent human communication? In natural conversation we are using ellipses, we are doing references, and to make the communication more effective and informative we are taking all these freedoms that you cannot codify as one or the other. You can't say, 'It means this.' You can only say, 'It means this in the context of this conversation.' So it's all context dependent. Vision can be more context independent. But spoken language more so cannot be context independent.¹⁵⁸

Both humor and speech, then, are context dependent. The meanings of humor and speech are highly dependent on where and how the information is conveyed. The combination of humor and speech may then be especially effective in obfuscating automated detection mechanisms, since current AI models are not well equipped to understand contextual clues.

That said, data scientist Christopher Penn acknowledged in his podcast that these limitations are changing with the latest advances in AI, especially multimodal models such as Microsoft's LLaVA and OpenAI's GPT4V. He explained that these models "can see and read. So they can take a text input and visual input and mix the two. It is not a stretch of the imagination to have a text model combined with an audio model, so that a machine can listen to that intonation" and understand different meanings based on tone of voice (Penn 2023).

The language in this quote is noteworthy in that it implies that the newest multimodal models have text and visual capabilities, but not audio capabilities. Audio may come next, but it currently lags behind equivalent technologies for text and images. This example demonstrates that latencies associated with the development of audio communication technologies persist even in the context of the most recent advancements. These latencies provide continuous opportunities in which speech can be used—often in combination with other obfuscation techniques, such as humor and uncommon languages—for open communication outside of the bounds of censorship and surveillance.

Regulatory Latency and the Audio-Visual Litany

In the remainder of this chapter, I suggest that this pattern—where the development of text and image AI models often precedes that of listening models—is indicative of a broader trend in the history and trajectory of AI for audio, particularly since the advent of deep learning in the 2010s. While especially pronounced since deep learning, latencies associated with the development of audio technologies can also be found in tracing the development of computer hardware. These audio latencies may also be connected to longstanding (though often contested) associations of vision with progress and modernity. By tracing these patterns and associations, I point toward how audio communication technologies and their regulatory frameworks often operate on a different time scale than visual or text technologies. Awareness of these differences

¹⁵⁸ Interview with author, November 15 2021.

can provide a deeper understanding of trends in the development of communication technologies and inform priorities for future development.

In a series of blog posts written in 2017-2018, engineer Daniel Rothmann noted differences in results for audio and visual data in AI. He stated, “We are doing machine vision to do machine hearing” (Rothmann 2017). Rothmann specifically discussed the context of an AI architecture called convolutional neural networks (CNNs). He explained that CNNs are based loosely on the human visual cortex and were initially designed for image processing, excelling in identifying patterns and localized features within visual data. Researchers later realized that CNNs could be applied to audio data as well if the audio data was represented as a spectrogram (Wyse 2017; Verma and Smith 2018).

A spectrogram is a visual representation of the spectrum of frequencies in a signal over time. It is useful to audio engineers because it represents audio data in convenient, compressed format. To make a spectrogram from a raw audio signal, a raw audio (wav) file is broken into samples; the standard is 44,100 samples per second. When compressed into a spectrogram, these 44,100 samples become a single data point on the time scale of a spectrogram. While this compression of data does lead to a loss of information, for most applications the amount of useful information in the 44,100 samples per second is quite sparse. Spectrograms provide a way to compress raw audio data so that more useful information is contained with less computer memory, and so that this information takes less computer power (and less time and money) to process.

Since spectrograms look like images, audio data contained in a spectrogram could be processed with CNNs, even though these CNNs were originally developed for images. Applying CNNs to audio was a convenient way to use an existing tool in a new context. However, Rothmann suggested that despite the apparent convenience and promise of applying image processing techniques to spectrograms, early applications of CNNs to audio were not “nearly as compelling as those...in the visual field” (Rothmann 2017). Tools developed first for images could be used for audio data, but the results of applying these tools to audio were, at least at first, somewhat disappointing.

Rothmann pointed out that this discrepancy might be due to the inherent differences in data representation between images and spectrograms. An image represents information locally: for example, in an image of a face all of the pixels representing an eye would be clustered together. A spectrogram, on the other hand, is a graph that represents information along axes: time on the horizontal axis, frequency on the vertical axis, and colors adding a third dimension to represent the amplitude (volume) of each frequency (see Figure 20). In image processing, a CNN might be trained to identify localized features, such as an eye in the top left corner of a face. However, in spectrograms valuable data is often contained in horizontal, parallel waveforms that span across the image. These horizontal waveforms can provide the grounds to distinguish one singer or instrument from another, for example. When applied to a spectrogram, a CNN originally developed for images might incorrectly focus on localized blobs, overlooking the critical information contained in the horizontal waveforms cutting across the spectrogram.

Furthermore, for the vertical axis in a spectrogram, shifting the frequency up or down could change how the sound is interpreted, since this would raise or lower the pitch. For images, the positioning along the x and y axes does not change the meaning of the image. Rothmann writes, “A face is still a face regardless of whether it is moved horizontally or vertically in an image.” In a spectrogram, however, “moving the frequencies of a male voice upwards could

change its meaning from man to child or goblin” (see Figure 21) (Rothmann 2018a). CNNs developed for images therefore originally were not optimized to identify aspects of audio data such as pitch and timbre. This fundamental difference in data structure and representation posed initial challenges when adapting image-based CNN techniques to audio data, necessitating further refinement and adaptation of these methods for effective audio processing.¹⁵⁹

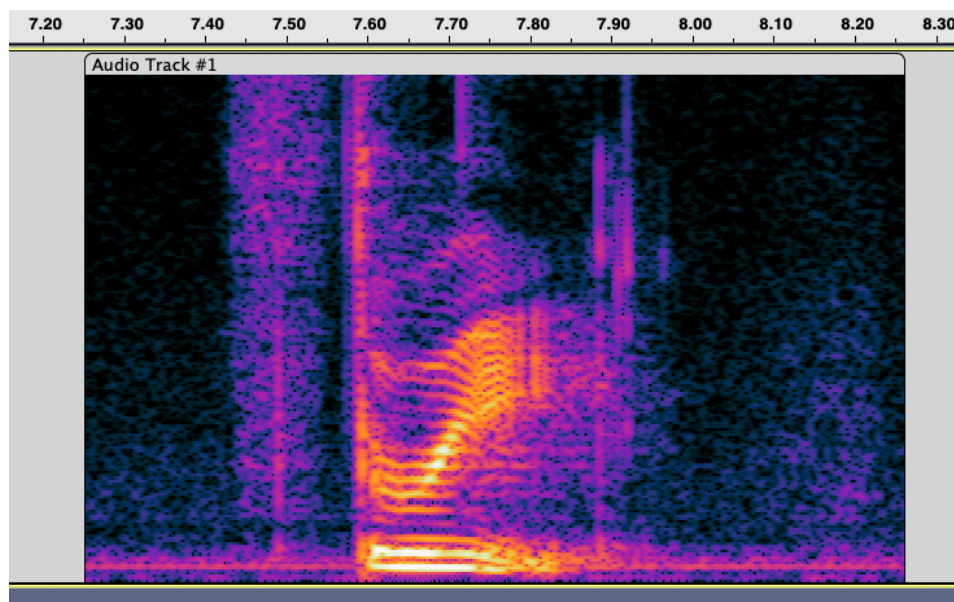


Figure 20. An example of a spectrogram. Time is represented on the horizontal axis, and frequency is represented on the vertical axis. Shading and color represent amplitude, or the volume at each frequency. Reproduced from <https://towardsdatascience.com/whats-wrong-with-spectrograms-and-cnns-for-audio-processing-311377d7ccd>

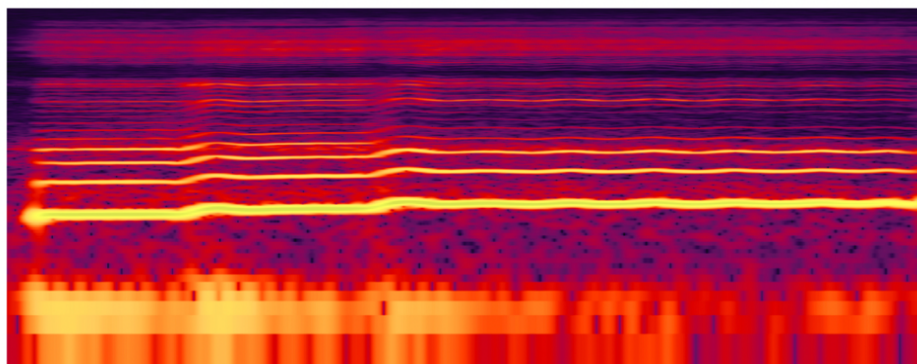


Figure 21. A spectrogram demonstrating non-localized representation of information. Here the frequencies track horizontally across the spectrogram—information that a CNN trained on images would have difficulty identifying. Reproduced from <https://towardsdatascience.com/whats-wrong-with-spectrograms-and-cnns-for-audio-processing-311377d7ccd>

¹⁵⁹ For examples of how CNNs can be adjusted to account for temporal and timbral information in music, see (Pons and Serra 2017; Pons et al. 2017).

CNNs were developed first for images and only later applied to audio, leading initially to disappointing results when this technology was first transferred to audio. This observation is especially significant because it is not an isolated example: since the shift to deep learning in the 2010s, there has been a trend that AI tools are quite often first developed in the context of visual or text data, and only later applied to audio. As another example, transformers are a groundbreaking deep learning model that are designed to weigh the importance of different elements in a sequence, capturing context and relationships more effectively. They were developed first for natural language processing and only later applied to images and audio processing. ZT explained that transformers are a logical application to text because they excel at capturing sequential patterns. It was then surprising that transformers were so successful in processing images, and that vision transformers could also be applied to speech. ZT said:

The transformer makes sense for text, because it tries to capture sequential patterns, but somehow they modify it [for images] to capture the pixels next to each other. [The pixels] also have some correlations. They tried to use the transformer to capture that information and were very successful. Then some models trained for vision were used in speech, and it's a vision transformer. It doesn't make any sense. Right now we cannot make a concrete conclusion on what model is best for which domain. So if some model is very accessible for one domain, the other domain will look at that model maybe they will use that in their own domain.¹⁶⁰

Both CNNs and transformers are representative of broader trends in the development of audio AI tools since deep learning. In fact, AI professionals I spoke with were hard-pressed to think of examples of AI applications since the age of deep learning that were derived for audio first.¹⁶¹

That said, while latencies in the development of AI tools for audio are especially prominent since deep learning, this trend has not always been so pronounced. Anumanchipalli explained that the different fields of AI research come from a “single source point” in the 1960s, with a “shared taxonomy of knowledge.” However, the field soon diverged into the separate domains of vision, graphics, computer architecture, and speech, each with its unique categorizations and representations. For example, visual processing focused on image features and object categorization, while speech processing dealt with the structured nature of language and its constraints.¹⁶²

Prior to deep learning, audio processing tools were largely developed within the realm of digital signal processing (DSP) rather than AI. DSP involves the use of mathematical and computational techniques to analyze, modify, and synthesize signals from audio, video, and other sensor data in a digital format. Various DSP techniques, such as hidden Markov models and the fast Fourier transform, were developed for or applied to audio and then later found applications

¹⁶⁰ Interview with author, November 15 2022.

¹⁶¹ Taiwanese AI researcher MT said, for example, “In recent years there are more and more advanced examples, like...dynamic convolutional networks, transformer models and attention mechanisms, and many other advanced techniques. Most of these very advanced models are not from audio. They either come from computer vision or from natural language processing” (Interview with author, June 17 2022).

¹⁶² Interview with author, November 15 2021.

beyond the audio field.¹⁶³ Even with broadly applicable DSP techniques coming out of audio processing, ZT explained that there was slow progress in speech recognition in part because this field required highly specialized knowledge in both statistics and acoustic engineering. However, the amount of expertise that was required changed with the onset of deep learning. Deep learning tools could now be broadly applied across domains, reducing the necessity for the specialized knowledge of acoustic engineers. As ZT recounted:

It's a very interesting story told by my advisor. When I was studying my PhD 20 years ago, my advisor told me that he was in Bell Labs. And he said that one supervisor in Bell Labs told him that every time he fired an acoustic expert—someone who is very good at acoustics but not an engineer—he will make big progress on ASR [automatic speech recognition]. So that's to say that the acoustic information is not important at that time. Engineering is more important. Recently another famous professor said that every time we fire an acoustic engineer we improve. So nowadays machine learning guys are important. Even if they don't know acoustic information, any engineer at all can make big progress.¹⁶⁴

As indicated here, with the advent of deep learning in the 2010s, which was enabled by improvements in computing hardware and increased availability of large amounts of data, the separate fields of AI once again converged, with neural networks developed in one domain more easily transferred to another.¹⁶⁵ Engineers are now able to process more data more quickly, leading to faster progress across domains including in speech and audio processing.

Even with this progress, the convergence of fields has created unanticipated challenges such as the loss of specialized domain knowledge, leading to claims such as Rothmann's that “we are doing machine vision to do machine listening.” Similarly, in the context of early music recommendation algorithms for platforms such as Spotify, anthropologist Nick Seaver observed that “how music sounds” was “surprisingly absent from most music recommenders” (Seaver 2022, 20). These platforms instead rely on tracing patterns in personal taste. Seaver writes, “Algorithmic recommendation has largely ignored the content of what it recommends, modeling

¹⁶³ For example, hidden Markov models (HMMs) were applied to speech recognition in the 1970s and later found wide use in bioinformatics (Jose et al. 2016). HMMs provided a statistical model for representing real-world data, which facilitated computational analysis and were foundational for AI (Mor et al. 2021). The fast Fourier transform (FFT), an early DSP technique, was initially proposed in 1965 and played a central role in the development of digital audio compression in the 1980s and 1990s (Sterne 2012, 7). FFTs have been widely applied in audio processing for speech and music analysis, and have also been applied to image processing and compression (Petrosian 2001).

¹⁶⁴ Interview with author, November 15 2022. Anumanchipalli provided a similar account: “So there is now a lot more shared taxonomy across fields, and that's due to the usage of similar infrastructure both in terms of the hardware and software that's powered by this deep learning era that we are currently in. That said, because of the niche that each of these fields exists in, it's hard to find experts that are equally adept at both of them. But I'd say that a researcher in one of these domains can very easily pick up a research paper in another domain and understand what's happening. Because there are analogs and correlates between both, it's just your input is different” (Interview with author, November 15 2021).

¹⁶⁵ Interview with author, November 15 2021. For discussion of the development of large scale datasets since the 2010s, see Orr and Crawford 2023.

taste as a pattern of interactions between users and items—which could just as easily be recipes or hotels as songs.”¹⁶⁶

This broader context since deep learning indicates that latencies in the development of AI audio tools have continued even with the most recent technologies. Since visual and text data often take fewer resources to gather and process (as McDowell observed), it typically makes sense to develop tools outside of the audio domain first and then later apply these tools to audio. This process can then lead to at least initially disappointing results in the audio domain. That said, what can be described as disappointing results also can indicate the continuing cycle of regulatory latency, which is not inherently good or bad. Recall that mainland Chinese internet users have relied on various obfuscation techniques to avoid censors: first homonyms, then embedding text in images, then using audio instead of text. Most recently, context-specific tools such as uncommon languages and humor combined with audio are especially effective. These obfuscation strategies constantly aim to be one step ahead of the latest regulatory technologies. Continuing latencies in the development of audio AI models since deep learning then imply that audio communication may continue to be especially beneficial in avoiding detection, even with the development of the newest, cutting-edge communication and regulatory tools.

While more pronounced since deep learning, the idea of computer technologies lagging in audio capabilities can be observed even before the widespread use of AI. A quick study of the history of computer hardware supports this point. Personal computers (PCs) marked a key turning point in the spread of computer technologies. In 1977, three PCs hit the market, known as the “1977 Trinity” and commonly considered the beginning of the PC industry. This trio included the Commodore PET, the TRS-80, and the Apple II. Of these, the Apple II was especially influential as one of the first fully assembled personal computers with color graphics and a user-friendly design. Even with these advancements in graphics, the Apple II did not have much in the way of audio capabilities. Apple II did have a speaker, but the speaker was limited to 1-bit output that could be switched on and off, producing only a clicking sound. People interested in listening to digitized audio would need an external sound card, but even sound cards were niche at the time. In fact, analog audio remained most prominent until the late 1990s, when developments in compression enabled more efficient digital storage of audio through the widespread use of the mp3.¹⁶⁷

In optimizing for graphics rather than audio, these early devices can perhaps be conceptualized as setting a precedent for the future development of computer hardware. This precedent is evident in more recent computer hardware trends. Deep learning, for example, is powered by graphics processing units (GPUs), which were originally designed for graphics applications such as videogames. The shift to using GPUs for deep learning in the early 2000s marked a significant breakthrough in the field. GPUs were originally designed to process the pixels in digitized images. This design also made them well-suited for the parallel processing required for matrix multiplication (the basis of deep learning algorithms), with the ability to perform thousands of calculations simultaneously. The highly parallel computations required for

¹⁶⁶ Seaver clarifies that “with the growth of budgets, data sets, and computational power, adding sound into the mix of signals used for recommendation has come to seem more plausible. Today, most of the large ensemble models used for music recommendation include, in some way, a representation of sound” (Seaver 2022, 97). This account indicates another example of latency in the context of audio. Recommendation algorithms—even those for music—first functioned without audio data, and only later began to increasingly incorporate audio.

¹⁶⁷ For a detailed history of audio compression, see Sterne 2012.

deep neural networks map perfectly onto GPUs, and the realization of this coincidence in the early 2000s led to significant progress in deep learning (Merritt 2023; Raina et al. 2009). That GPUs are the foundation for deep learning also means that at a fundamental level, the very hardware that enables deep learning was originally invented for processing visual information, rather than audio data.

While the adoption of GPUs for deep learning can be explained as a technical convenience, it may also suggest the extent to which visuality is deeply ingrained in the way that deep learning and computer technologies in general function. This phenomenon in fact evokes the concept that Jonathan Sterne calls the “audio-visual litany,” or the persistent dichotomy that positions sound and vision in opposition to one another. Sterne writes that the juxtaposition of audio and vision are ideologically loaded, going back to Judeo-Christian associations of hearing with the divine, pureness, and spirituality (a “kind of pure interiority”), and vision as intellectual and analytical. The concept of the audio-visual litany aims to reveal this dichotomy, tracing it to Christian spiritualism reaching back to Plato, and more recently in the work of Walter Ong. Though Ong’s work was influential in bringing sound as a method for understanding the world to the attention of many scholars, his emphasis on sound is rooted in a problematic binary in which hearing is considered spiritual and close to God, while vision suggests disengaged, objective analysis. In revealing the presence of the audio-visual litany, Sterne hoped to “provide an alternative to the pervasive narrative that says that, in becoming modern, Western culture moved away from a culture of hearing to a culture of seeing” (Sterne 2003, 2).¹⁶⁸

Progress in visual applications of AI and the relative latency of audio AI developments may then suggest the persistence of the audio-visual litany as deeply ingrained in the societal and cultural forces driving technological development. While there are certainly many technical and practical forces that influence the development of AI technologies (such as data processing time, computing power resources, and data availability, as previously discussed), some of these factors may in fact stretch back to the earliest computer hardware and even before, shaped to some extent by largely internalized, subconscious associations of computer technologies and visual perception with progress and modernity.

A discussion of the audio-visual litany in the context of the development of AI then invites creative re-imaginings of current technologies. For example, after noting the early disappointing results of CNNs for audio, Rothmann went on to suggest the development of an experimental algorithm based on human hearing (Rothmann 2018b). Thinking through the audio-visual litany in the context of the development of AI then invites questions such as: What would be the current state of computer technologies if the first PCs had included advanced audio capabilities rather than graphics? How might audio processing be different today if deep learning was based on sound cards, rather than GPUs? How might AI models that are developed first for audio be widely applied across domains? While these questions may sound farfetched in the current context of computer technologies, they encourage a more critical understanding of the

¹⁶⁸ The task of challenging the problematic dichotomy between hearing and vision has been taken up by various scholars across interdisciplinary fields. Much of the field of sound studies is concerned with this topic. In one especially influential example, Ana Maria Ochoa-Guatier studies the archive from an aural perspective by revealing what can be heard within written sources from nineteenth century Colombia. She emphasizes the ways in which sound was used to establish and maintain social hierarchies as well as to challenge them, demonstrating the extent to which sound and aural understandings have been overlooked in the telling of history (Ochoa Gautier 2014).

various social, cultural, and technical forces that shape the current state of communication and its regulation.

Conclusion

This chapter has drawn out a continuing tension: that of the race toward progress and innovation as opposed to the drive toward further control. Within these juxtaposed processes, the cycle of regulatory latency is a continuous presence, revealing a pattern in which the widespread use of communication technologies often precedes the development and implementation of corresponding regulatory technologies. This latency then opens a window for unregulated communication.

While there are risks associated with this unregulated communication, as in the case of the unregulated spread of disinformation, windows of regulatory latency can also create opportunities for open communication, especially in places with highly regulated media environments such as mainland China. In the context of China's internet, people employ various techniques for taking advantage of regulatory latencies. These techniques often leverage delays or shortcomings in technological developments. Latencies in the collection of diverse speech data, for instance, allow people to communicate more openly with less common languages, dialects, and accents. The lack of contextual understanding in current AI models means that people can use tools such as humor and satire in combination with speech to circumvent automated detection tools. A pattern in these obfuscation techniques is the leveraging of regulatory latencies specifically for voice communication. Speech contains a high degree of variation and diversity (accents, dialects), and meaning conveyed through speech is highly dependent on context and tone of voice. At the same time, due to deeply embedded technical and cultural reasons, data collection and processing for speech and audio often trail behind those for text or images.

Latency at face value indicates a lag or a gap—concepts that typically have negative connotations amidst the constant drive for progress, innovation, and regulation. At the same time, latency in the context of cross-strait communication can indicate an opportunity for open communication. In this sense, latency disrupts the constant drive toward progress and innovation, as well as the constant drive toward further control. Latency demonstrates that delays, lags, and gaps are not inherently positive or negative, revealing the unexpected benefits of taking things slowly.

Works Cited

- Abe, Marie. 2018. *Resonances of Chindon-Ya*. Wesleyan University Press.
- Ahern, Deirdre. 2021. "Regulatory Lag, Regulatory Friction and Regulatory Transition as FinTech Disenablers: Calibrating an EU Response to the Regulatory Sandbox Phenomenon." *European Business Organization Law Review* 22 (3): 395–432.
- Allan, Michael. 2019. "Old Media / New Futures: Revolutionary Reverberations of Fanon's Radio." *PMLA* 134 (1): 188–93.
- Allen, Darcy, and Chris Berg. 2017. "Regulation and Technological Change." *Australia's Red Tape Crisis*, 218–30.
- Anderson, Benedict. 2006. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Verso.
- Andrews, Scott. 2014. "WoW Archivist: WoW in China, an Uncensored History." Engadget. 2014.
- Associated Press. 2022. "Taiwan's Compulsory Military Service Is Extended to a Year :." *NPR*, December 27, 2022.
- Barrington, Sarah, Romit Barua, Gautham Koorma, and Hany Farid. 2023. "Single and Multi-Speaker Cloned Voice Detection: From Perceptual to Learned Features." *WIFS 2023 - IEEE Workshop on Information Forensics and Security*, 1–6.
- Baucom, Ian. 2001. "Frantz Fanon's Radio: Solidarity, Diaspora, and the Tactics of Listening." *Contemporary Literature* 42 (1): 15.
- BBC News*. 2020. "Zoom Banned by Taiwan's Government over China Security Fears." April 7, 2020.
- Benkler, Yochai. 2006. *The Wealth of Networks : How Social Production Transforms Markets and Freedom*. Yale University Press.
- Bessire, Lucas. 2012. "'We Go Above': Media Metaphysics and Making Moral Life on Ayoreo Two-Way Radio." In *Radio Fields : Anthropology and Wireless Sound in the 21st Century*, edited by Daniel Fisher and Lucas Bessire, 197–214. New York University Press.
- Bessire, Lucas, and Daniel Fisher. 2012. "Introduction: Radio Fields." In *Radio Fields: Anthropology and Wireless Sound in the 21st Century*, edited by Lucas Bessire and Daniel Fisher, 1–47. New York University Press.
- Biselli, Anna. 2018. "BAMF Software Puts People in Danger." *Vice*, August 18, 2020.
- Bogost, Ian. 2011. *How to Do Things With Videogames*. University of Minnesota Press.
- . 2015. *How to Talk about Videogames*. University of Minnesota Press.
- Bolter, Jay David, and Richard Grusin. 2016. *Understanding New Media*. SAGE Publications.
- Born, Georgina. 2013. *Music, Sound and Space: Transformations of Public and Private*

- Experience*. Cambridge University Press.
- Bozkurt, Ilker Nadi, Anthony Aguirre, Balakrishnan Chandrasekaran, P. Brighten Godfrey, Gregory Laughlin, Bruce Maggs, and Ankit Singla. 2017. "Why Is the Internet so Slow?!" In *Lecture Notes in Computer Science*, 10176:173–87. Springer Verlag.
- Browne, Craig, and Paula Diehl. 2019. "Conceptualising the Political Imaginary: An Introduction to the Special Issue." *Social Epistemology* 33 (5): 393–97.
- Buckley, Chris, and Amy Chang Chien. 2024. "Experts See a Message in Chinese Balloons Flying Over Taiwan." *The New York Times*, January 4, 2024.
- Chai, Winberg. 2000. "The Transformation of the Mass Media in Taiwan since 1950: Introduction." *Asian Affairs* 27 (3): 133–40.
- Chen, Chien Min, and Tsung Hsien Tsai. 2019. "Tourist Motivations in Relation to a Battlefield: A Case Study of Kinmen." *Tourism Geographies* 21 (1): 78–101.
- Chen, Chu-chu 陳楚楚. 2020. "Why Did 'Chairman Rabbit,' the Third-Generation Red KOL, Speak about Hong Kong and Became Popular in the Simplified Chinese Public Opinion Circle? 紅三代KOL兔主席的香港論述, 為何在簡體輿論圈走紅?" *Initium Media*, November 16, 2020.
- Chen, Jianqing. 2023. "Bullet-Titling in Twenty-First Century China: Bullet-like Tacitility, Soft Touch and Diachronic Simultaneity." *Journal of Visual Culture* 22 (1): 3–24.
- Chen, Kuan-Hsing. 2010. *Asia as Method*. Duke University Press.
- Chen, Lai Chi. 2009. "The Value Chain In the Asian Online Gaming Industry: A Case Study of Taiwan." University of Westminster.
- Chen, Liya 陳莉雅. 2021. "Clubhouse 「兩岸房」三日記: 如果沒有牆, 我們會跟彼此聊什麼? ." *Initium Media 端傳媒*. 2021.
- Cheng, Chen Ching. 2019. "'The Eternal Sweetheart for the Nation': A Political Epitaph for Teresa Teng's Music Journey in Taiwan." In *Made in Taiwan: Studies in Popular Music*, edited by Eva Tsai, Tung-Hung Ho, and Miaoju Jian, 201–10. Taylor and Francis.
- Chhillar, Deepika, and Ruth V. Aguilera. 2022. "An Eye for Artificial Intelligence: Insights Into the Governance of Artificial Intelligence and Vision for Future Research." *Business and Society* 61 (5): 1197–1241.
- China World Group, World Journal*. 2021. "You Lai Suan Zhong Huang Mingzhi Tong Zhongguo Geshou Hechang `qiang Wai' Kan Ku Bai Wan Fensi [Souring Again towards China, Huang Mingzhi Collaborates with Chinese Singers to Sing 'The Wall,' Moving Millions of Fans to Tears. 又來酸中 黃明志同中國歌手合唱「牆外」看哭百萬粉絲]," November 15, 2021.
- Clifford, James. 1986. "Introduction: Partial Truths." In *Writing Culture: The Poetics and Politics of Ethnography*, edited by James Clifford and George E. Marcus, 1–26. Berkeley: University of California Press.

- Connor, Steven. 1997. "The Modern Auditory I." In *Rewriting the Self: Histories From the Renaissance to the Present*, edited by Roy Porter, 203–23.
- Couldry, Nick. 2008. "Media and the Problem of Voice." In *Participation and Media Production: Critical Reflections on Content Creation.*, edited by Nico Carpentier and Benjamin Cleen, 15–26. Cambridge Scholars Publishing.
- Cronin, Richard. 2022. "Semiconductors and Taiwan's 'Silicon Shield.'" *Stimson Center Issue Brief: Asia and Indo-Pacific*.
- Dahlgren, Peter. 2005. "The Internet, Public Spheres, and Political Communication: Dispersion and Deliberation." *Political Communication* 22 (2): 147–62.
- Davies, Hugh. 2022. "The Revolution Will Not Be Gamified: Videogame Activism and Playful Resistance across the Sinophone." *British Journal of Chinese Studies* 12 (2).
- DeCamp, Whitney, and Christopher J. Ferguson. 2016. "The Impact of Degree of Exposure to Violent Video Games, Family Background, and Other Factors on Youth Violence." *Journal of Youth and Adolescence* 46 (2): 388–400.
- Devine, Kyle, and Alexandrine Boudreault-Fournier, eds. 2021. *Audible Infrastructures*. Oxford University Press.
- Diaz Ruiz, Carlos, and Tomas Nilsson. 2023. "Disinformation and Echo Chambers: How Disinformation Circulates on Social Media Through Identity-Driven Controversies." *Journal of Public Policy and Marketing* 42 (1): 18–35.
- Doublethink Lab. 2024. "2024 Taiwan Elections: Foreign Influence Observation—Preliminary Statement," February 27, 2024.
- Eidsheim, Nina Sun. 2019. *The Race of Sound: Listening, Timbre, and Vocality in African American Music*. Duke University Press.
- Election Study Center, National Chengchi University. 2023. "Taiwanese / Chinese Identity (1992/06-2023/06)."
- Erlmann, Veit. 2004. *Hearing Cultures: Essays on Sound, Listening, and Modernity*. Berg.
- Fanon, Frantz; and Haakon Chevalier (trans). 1965. "This Is the Voice of Algeria." In *A Dying Colonialism (Orig. 1959)*. Monthly Review Press.
- Federer, Julia Palmiano. 2021. "Toward a Normative Turn in Track Two Diplomacy? A Review of the Literature." *Negotiation Journal* 37 (4): 427–50.
- Feld, Steven. 2015. "Listening to Histories of Listening: Collaborative Experiments in Acoustemology with Nii Otoo Annan." In *Musical Listening in the Age of Technological Reproduction*, edited by Gianmario Borio, 92-103. Routledge.
- — —. 2017. "On Post-Ethnomusicology Alternatives: Acoustemology." In *Perspectives on a 21st Century Comparative Musicology: Ethnomusicology or Trans-Cultural Musicology?*, edited by Giovanni Giuriati.
- Feld, Steven, and Keith H. Basso. 1996. *Senses of Place*. University of Washington Press.

- Fisher, Daniel. 2012. "From the Studio to the Street: Producing the Voice in Indigenous Australia." In *Radio Fields : Anthropology and Wireless Sound in the 21st Century*, edited by Daniel Fisher and Lucas Bessire, 68–88.
- Frith, Simon. 1988. "Why Do Songs Have Words?" In *Music for Pleasure*, 105–28. New York: Routledge.
- Garrett, Daniel, and Wing-chung Ho. 2014. "Hong Kong at the Brink: Emerging Forms of Political Participation in the New Social Movement." In *New Trends of Political Participation in Hong Kong*, edited by Joseph Y.S. Cheng. City University of Hong Kong Press.
- Gibbs, Martin, Greg Wadley, and Peter Benda. 2006. "Proximity-Based Chat in a First Person Shooter: Using a Novel Voice Communication System for Online Play." *Proceedings of the 3rd Australasian Conference on Interactive Entertainment*, no. December: 96–102.
- Gimmler, Antje. 2001. "Deliberative Democracy, the Public Sphere and the Internet." *Philosophy & Social Criticism* 27 (4): 21–39.
- Ginsburg, Faye. 2012. "Radio Fields: An Afterword." In *Radio Fields : Anthropology and Wireless Sound in the 21st Century*, edited by Daniel Fisher and Lucas Bessire, 269–78. New York University Press.
- Goh, Brenda. 2021. "Clubhouse App Blocked in China, Added to 'Great Firewall.'" Reuters. February 9, 2021.
- Gold, Thomas B. 1986. *State and Society in the Taiwan Miracle*. Routledge.
- Gradinaru, Camelia. 2018. "Small Talk in Our Digital Everyday Life: The Contours of a Phatic Culture." *Meta* 10 (2): 459–72.
- Graham-Harrison, Emma, and Chi-Hui Lin. 2023. "'Into Brain and the Heart': How China Is Using Apps to Woo Taiwan's Teenagers." *The Guardian*, August 3, 2023.
- Green, Michael, Kathleen Hicks, Zack Cooper, John Schaus, and Jake Douglas. 2017. "Countering Coercion in Maritime Asia: The Theory and Practice of Gray Zone Deterrence." Center for Strategic and International Studies.
- Greene, J Megan. 2008. *The Origins of the Developmental State in Taiwan Science Policy and the Quest for Modernization*. Harvard University Press.
- Griffith, Erin, and Taylor Lorenz. 2020. "The Hot New Thing in Clubby Silicon Valley? An App Called Clubhouse ." *The New York Times*, May 19, 2020.
- Griffiths, James T. 2019. "The Great Firewall of China: How to Build and Control an Alternative Version of the Internet." Zed Books.
- Guilbault, Jocelyne. 2019. "Keynote: Party Music, Affect and the Politics of Modernity." *Journal of World Popular Music* 6 (2): 173–92.
- Guy, Nancy. 2019. *Garbage Truck Music and Sustainability in Contemporary Taiwan: From Cockroaches to Beethoven and Beyond*. University of Illinois Press.

- . 2021. “Listening to Taiwan’s Musical Garbage Trucks: Hearing the Slow Violence of Environmental Degradation.” In *Resounding Taiwan: Musical Reverberations Across a Vibrant Island*, edited by Nancy Guy, 180–96, Taylor and Francis.
- Habermas, Jurgen. 1991. *The Structural Transformation of the Public Sphere*. The MIT Press.
- Han, Rongbin. 2018. *Contesting Cyberspace in China : Online Expression and Authoritarian Resilience*. Columbia University Press.
- Hartono, M. Paulina. 2020. “‘A Good Communist Style’: Sounding Like a Communist in Twentieth Century China.” *Representations* 151 (1): 26–50.
- Ho, Ming Sho. 2022. “Changing Memory of the Tiananmen Incident in Taiwan: From Patriotism to Universal Values (1989–2019).” *China Information* 36 (1): 90–111.
- Hochheimer, John L. 1993. “Organizing Democratic Radio: Issues in Praxis.” *Media, Culture & Society* 15: 473–86.
- Hsieh, Jennifer C. 2021. “Noisy Co-Existence: Contestations of Renao and Zaoyin amidst Taiwan’s Noise Control System.” *Resounding Taiwan: Musical Reverberations Across a Vibrant Island*, edited by Nancy Guy, 165–79, Taylor and Francis.
- Hsieh, Jennifer C. 2017. “Noise Governance and the Hearing Subject in Urban Taiwan.” Stanford University.
- . 2019. “Piano Transductions: Music, Sound and Noise in Urban Taiwan.” *Sound Studies* 5 (1): 4–21.
- Hsu, Eric. 2024. “2024 Taiwan Election: The Increasing Polarization of Taiwanese Politics - Reinforcement of Conspiracy Narratives and Cognitive Biases.” *Doublethink Lab*, April 8, 2024.
- Huang, Zong-ding 黃宗鼎. 2022. “Research on Taiwan’s Psychological Warfare Against Mainland China—Centered on RTI (1950-1998)央廣心戰檔案研究報告: 臺灣對中國大陸之心戰研究 -以央廣為中心 (1950-1998) .”
- Human Rights Watch*. 2017. “China: Voice Biometric Collection Threatens Privacy,” October 22, 2017.
- Hung, William. 2024. “Media Literacy Education: Taiwan’s Key to Combating Disinformation.” *Global Taiwan Brief* 9 (5).
- Hvistendahl, Mara. 2020. “How a Chinese AI Giant Made Chatting—and Surveillance—Easy.” *Wired*, May 18, 2020.
- “Instagram Appears Blocked in China.” 2014. BBC News. September 29, 2014.
- Inwood, Heather. 2022. “Towards Sinophone Game Studies.” *British Journal of Chinese Studies* 12 (2).
- Jack, Robert H., Adib Mehrabi, Tony Stockman, and Andrew McPherson. 2018. “Action-Sound Latency and the Perceived Quality of Digital Musical Instruments: Comparing Professional Percussionists and Amateur Musicians.” *Music Perception* 36 (1): 109–28.

- James, Paul. 2019. "The Social Imaginary in Theory and Practice." In *Revisiting the Global Imaginary: Theories, Ideologies, Subjectivities: Essays in Honor of Manfred Steger*, edited by C. Hudson and E. K. Wilson. Palgrave Macmillan.
- Jones, Andrew F. 2020. *Circuit Listening: Chinese Popular Music in the Global 1960s*. Minneapolis: Univ Of Minnesota Press.
- Jose, Shary, Prashant Nair, Vinai George Biju, Blessy Baby Mathew, and C. M. Prashanth. 2016. "Hidden Markov Model: Application towards Genomic Analysis." *2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT)*, August.
- Joseph, George, and Debbie Nathan. 2019. "Prisons Across the US Are Quietly Building Databases of Incarcerated People's Voice Prints." *The Intercept*, January 30, 2019.
- Kadar, Daniel Z., Michael Haugh, and Wei Lin Melody Chang. 2013. "Aggression and Perceived National Face Threats in Mainland Chinese and Taiwanese CMC Discussion Boards." *Multilingua* 32 (3): 343–72.
- Kakar, Akshaan. 2022. "Making the Hallway More Relevant with Machine Learning." Clubhouse Blog.
- Ke, Shun-Chih. 2000. "The Emergence, Transformation, and Disintegration of Alternative Radio in Taiwan: From Underground Radio to Community Radio." *Journal of Communication Inquiry* 24 (4): 412–29.
- Keogh, Brendan. 2018. *A Play of Bodies: How We Perceive Videogames*. MIT Press.
- King, Gary, Jennifer Pan, and Margaret E. Roberts. 2017. "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument." *American Political Science Review* 111 (3): 484–501.
- Kittler, Friedrich. 1999. *Gramophone, Film, Typewriter (Orig. 1987)*. Stanford University Press.
- Kloet, Jeroen de, Thomas Poell, Zeng Guohua, and Chow Yiu Fai. 2019. "The Platformization of Chinese Society: Infrastructure, Governance, and Practice." *Chinese Journal of Communication* 12 (3): 249–56.
- Kuo, Kaiser. 2021. "The Xinjiang Camps on Clubhouse." *Sinica Podcast by SupChina*, February 11, 2021.
- LaBelle, Brandon. 2018. "Sonic Agency," 179.
- Lacey, Kate. 2013. *Listening Publics: The Politics and Experience of Listening in the Media Age*. Wiley.
- — —. 2023. "The Labour of Listening in Troubled Times." *Journal of Sonic Studies*.
- Ladden-Hall, Dan. 2024. "Google Explains Why Its AI Tool Told Users to Eat Rocks and Glue Pizzas." *The Daily Beast*, May 31, 2024.
- Larkin, Brian. 2006. *Signal and Noise: Media, Infrastructure, and Urban Culture in Nigeria*. Duke University Press.

- Larson, David B., Hugh Harvey, Daniel L. Rubin, Neville Irani, Justin R. Tse, and Curtis P. Langlotz. 2021. "Regulatory Frameworks for Development and Evaluation of Artificial Intelligence–Based Diagnostic Imaging Algorithms: Summary and Recommendations." *Journal of the American College of Radiology* 18 (3): 413–24.
- Lefebvre, Henri. 1992. *Rhythmanalysis: Space, Time and Everyday Life*. Edited by Gerald Moore and Stuart Elden. Continuum.
- Leffer, Lauren. 2024. "AI Audio Deepfakes Are Quickly Outpacing Detection." *Scientific American*, January 26, 2024.
- Lei, Wei. 2019. *Radio and Social Transformation in China*. Routledge.
- Lekner, Dayton. 2022. "States of Diffusion: Ideology, Text, Voice, and Sound in Cold War Chinas." *Modern China* 49 (2): 1–33.
- Li, Jie. 2020. "Revolutionary Echoes: Radios and Loudspeakers in the Mao Era." *Twentieth-Century China* 45 (1): 25–45.
- . 2022. "The Hot Noise of Open-Air Cinema." *Sensing China: Modern Transformations of Sensory Culture*, 199–228.
- Li, Jinying. 2017. "The Interface Affect of a Contact Zone: Danmaku on Video-Streaming Platforms." *Asiascape: Digital Asia* 4 (3): 233–56.
- Li, Laura, and Scott Gregory. 2009. "The Central Broadcasting System at 80." *Taiwan Panorama*.
- Li, Wei-ping. 2023. "Seeing Is Not Believing—Deepfakes and Cheap Fakes Spread during the 2024 Presidential Election in Taiwan." *Taiwan FactCheck Center*, December 25, 2023.
- . 2024. "Seeing Is Not Believing (Part II) - AI Videos Spread during the 2024 Presidential Election in Taiwan." *Taiwan FactCheck Center*, February 19, 2024.
- Liao, Ping-Hui. 1999. "Postcolonial Studies in Taiwan: Issues in Critical Debates." *Postcolonial Studies* 2 (2): 199–211.
- Lin, Belle. 2018. "Amazon's Accent Recognition Technology Could Tell the Government Where You're From." *The Intercept*, November 15, 2018.
- Lin, Chih Yuan, and Mateus Lee. 2020. "Taiwan's Opening Policy to Chinese Tourists and Cross-Strait Relations: The Impacts on Inbound Tourism into Taiwan." *Tourism Economics* 26 (1): 27–44.
- Lin, Holin, and Chuen-Tsai Sun. 2011. "A Chinese Cyber Diaspora: Contact and Identity Negotiation on Taiwanese WoW Servers." In *Proceedings of DiGRA 2011 Conference: Think Design Play*.
- . 2019. "How Video Game Techno-Governance Shapes Social Interaction." *Delphi Interdisciplinary Review of Emerging Technologies* 2 (1): 48–51.
- Lin, Judy. 2024. "TSMC Dominates Global Foundry Market with 61 Percent Market Share in 4Q23, Says Counterpoint." *Digittmes Asia*, May 3, 2024.

- Lin, Mei-Hua 林美華. 2009. “傾聽戰地的聲音：金門的戰地廣播（1949-1992） Listening to the Sounds of a Battlefield: Broadcasting in Wartime Kinmen（1949-1992）.” National Kinmen University of Technology 國立金門技術學院.
- Lo, William Yat Wai. 2020. “China Suspends Study in Taiwan over Cross-Strait Relations.” *University World News*. 2020.
- Long, Qiao, Chingman, Gigi Lee, and Luisetta Mudie. 2021. “China Clamps Down on Software Used to Disguise Voiceprints.” *Radio Free Asia*, March 18, 2021.
- Lu, Zhicong, Chenxinran Shen, and Jiannan Li. 2021. “More Kawaii than a Real-Person Live Streamer: Understanding How the Otaku Community Engages with and Perceives Virtual Youtubers.” *Conference on Human Factors in Computing Systems - Proceedings*.
- Mainland Affairs Council, Republic of China (Taiwan). “Statistics on Mainland Students Researching and Studying for Degrees in Taiwan.” *Mainland Chinese Students in Taiwan*. https://www.mac.gov.tw/en/News_Content.aspx?n=CA7B5FA9C0EC7005&s=54D121541C911FB1&sms=D645444CA321A4FA&Create=1
- Maizland, Lindsay. 2022. “China’s Repression of Uyghurs in Xinjiang.” *Council on Foreign Relations*.
- Malinowski, Bronislaw. 1923. “The Problem of Meaning in Primitive Languages.” In *The Meaning of Meaning*, edited by C. K. Ogden and I. A. Richards, 146–52. Routledge.
- Markey, Patrick M., and Christopher J. Ferguson. 2017. *Moral Combat: Why the War on Violent Video Games Is Wrong*. BenBella Books.
- Markey, Patrick M., Charlotte N. Markey, and Juliana E. French. 2015. “Violent Video Games and Real-World Violence: Rhetoric versus Data.” *Psychology of Popular Media Culture* 4 (4): 277–95.
- Mazarr, Michael. 2015. *Mastering the Gray Zone: Understanding a Changing Era of Conflict*. United States Army War College Press.
- McCombs, Maxwell E., and Donald L. Shaw. 1972. “The Agenda-Setting Function of Mass Media.” *The Public Opinion Quarterly* 36 (2): 176–87.
- McEnaney, Tom. 2019. “This American Voice: The Odd Timbre of a New Standard in Public Radio.” In *The Oxford Handbook of Voice Studies*, edited by Nina Eidsheim; Katherine Meizel. Oxford University Press.
- McLuhan, Marshall. 1994. *Understanding Media (Orig. 1964)*. MIT Press.
- McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. “Birds of a Feather: Homophily in Social Networks.” *Annual Review of Sociology* 27 (Volume 27, 2001): 415–44.
- Merritt, Rick. 2023. “Why GPUs Are Great for AI.” *NVIDIA Blog*, December 4, 2023.
- Miller, Chris. 2022. *Chip War: The Fight for the World’s Most Critical Technology*. Simon & Schuster.

- Miller, Edythe S. 1996. "Economic Regulation and New Technology in the Telecommunications Industry." *Journal of Economic Issues* 30 (3): 719–35.
- Mina, An Xiao. 2014. "Batman, Pandaman and the Blind Man: A Case Study in Social Change Memes and Internet Censorship in China." *Journal of Visual Culture* 13 (3): 359–75.
- . 2018. *Memes to Movements : How the World's Most Viral Media Is Changing Social Protest and Power*. Beacon Press Books.
- Minssen, Timo, Sara Gerke, Mateo Aboy, Nicholson Price, and Glenn Cohen. 2020. "Regulatory Responses to Medical Machine Learning." *Journal of Law and the Biosciences* 7 (1): 1–18.
- Mor, Bhavya, Sunita Garhwal, and Ajay Kumar. 2021. "A Systematic Review of Hidden Markov Models and Their Applications." *Archives of Computational Methods in Engineering* 28 (3): 1429–48.
- Moskowitz, Marc L. 2009. *Cries of Joy, Songs of Sorrow: Chinese Pop Music and Its Cultural Connotations*. Honolulu: University of Hawaii Press.
- Mouffe, Chantal. 2005. *On the Political*. Routledge.
- Mozur, Paul, Amy Chang Chien, and Michael D. Shear. 2022. "Nancy Pelosi Arrives in Taiwan, Drawing a Sharp Response From Beijing." *The New York Times*, August 2, 2022.
- Muñoz, José Esteban. 1999. *Disidentifications: Queers of Color and the Performance of Politics*. University of Minnesota Press.
- Nahon, Karine, and Jeff Hemsley. 2013. *Going Viral*. Polity Press.
- Newland, Sara A. 2023. "Paradiplomacy as a Response to International Isolation: The Case of Taiwan." *Pacific Review* 36 (4): 784–812.
- Newton, Casey. 2021. "Elon Musk Just Showed How Clubhouse Can Succeed." *The Verge*, February 1, 2021.
- Ng, Jason Q. 2013. *Blocked on Weibo : What Gets Suppressed on China's Version of Twitter (and Why)*. New Press.
- Ngai, Sianne. 2005. "The Cuteness of the Avant-Garde." *Critical Inquiry* 31 (4): 811–47.
- . 2012. *Our Aesthetic Categories*. Harvard University Press.
- Nissenbaum, Helen, and Finn Brunton. 2015. *Obfuscation: A User's Guide for Privacy and Protest*. MIT Press.
- Ochoa Gautier, Ana Maria. 2014. *Aurality: Listening and Knowledge in Nineteenth-Century Colombia*. Duke University Press.
- OECD. 2021. *Development Co-Operation Report 2021: Shaping a Just Digital Transformation*. OECD Publishing.
- Orr, Will, and Kate Crawford. 2023. "The Social Construction of Datasets: On the Practices, Processes and Challenges of Dataset Creation for Machine Learning." *New Media & Society (forthcoming)*.

- Owens, Golden Marie. 2023. “‘Hey Google, Talk Like Issa’: Black Voiced Digital Assistants and the Reshaping of Racial Labor.” *Sounding Out!*, June 5, 2023.
- Papacharissi, Zizi. 2002. “The Virtual Sphere.” *New Media & Society* 4 (1): 9–27.
- Paris, Britt, and Joan Donovan. 2019. “Deepfakes and Cheap Fakes.” *Data & Society*, 47.
- Pariser, Eli. 2011. *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. Penguin.
- Peng, Chun-Yi. 2021. *Mediatized Taiwanese Mandarin*. Singapore: Springer Singapore.
- Penn, Christopher S. 2023. “Mind Readings: Why AI Struggles With Sarcasm,” October 24, 2023.
- Perry, Elizabeth J. 2013. “Moving The Masses: Emotion Work In The Chinese Revolution.” *Mobilization* 7 (2): 111–28.
- Peters, John Durham. 2015. *The Marvelous Clouds*. Chicago University Press.
- Petrosian, Arthur. 2001. “Transform Coding of Signals with Bounded Finite Differences: From Fourier to Walsh, to Wavelets.” In *Wavelets in Signal and Image Analysis. Computational Imaging and Vision*, edited by A.A. Petrosian and F.G. Meyer, 245–80. Dordrecht: Springer.
- Pfeifer, Michelle. 2023. “Your Voice Is (Not) Your Passport.” *Sounding Out!*, June 12, 2023.
- Pons, Jordi, and Xavier Serra. 2017. “Designing Efficient Architectures for Modeling Temporal Features with Convolutional Neural Networks.” In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*.
- Pons, Jordi, Olga Slizovskaia, Rong Gong, Emilia Gómez, and Xavier Serra. 2017. “Timbre Analysis of Music Audio Signals with Convolutional Neural Networks.” In *25th European Signal Processing Conference, EUSIPCO 2017*.
- Price, Gareth. 2019. *Language, Society, and the State: From Colonization to Globalization in Taiwan*. De Gruyter.
- Raina, Rajat, Anand Madhavan, and Andrew Y. Ng. 2009. “Large-Scale Deep Unsupervised Learning Using Graphics Processors.” *ACM International Conference Proceeding Series*.
- Raoufi, Mohsen, Heiko Hamann, and Pawel Romanczuk. 2024. “Messengers: Breaking Echo Chambers in Collective Opinion Dynamics with Homophily.” *ArXiv*, 1–28.
- Rawnsley, Gary D. 1996. *Radio Diplomacy and Propaganda*. St. Martin’s Press.
- . 2000. *Taiwan’s Informal Diplomacy and Propaganda*. Palgrave Macmillan UK.
- Remnick, David. 2021. “Clubhouse Opens a Window for Free Expression in China.” *The New Yorker*. 2021.
- Rheingold, Howard. 2003. *Smart Mobs : The Next Social Revolution*. Perseus Pub.
- Rigger, Shelley. 2011. *Why Taiwan Matters: Small Island, Global Powerhouse*. Lanham, Md: Rowman & Littlefield.

- — —. 2021. *The Tiger Leading the Dragon: How Taiwan Propelled China's Economic Rise*. Rowman & Littlefield.
- Roberts, Margaret E. 2017. *Censored: Distraction and Diversion Inside China's Great Firewall*. Taylor and Francis.
- Roth, Emma. 2021. "The Global Version of Steam Appears to Be Banned in China." *The Verge*, December 25, 2021.
- Rothmann, Daniel. 2017. "The Promise of AI in Audio Processing." *Towards Data Science*, December 26, 2017.
- — —. 2018a. "What's Wrong with CNNs and Spectrograms for Audio Processing?" *Towards Data Science*, March 25, 2018.
- — —. 2018b. "Human-Like Machine Hearing With AI (1/3)." *Towards Data Science*, June 5, 2018.
- Rowen, Ian. 2023. *One China, Many Taiwans: The Geopolitics of Cross-Strait Tourism*. Ithaca: Cornell University Press.
- Sato, Tomoka. 2022. "Linguistic Assimilation in China: Past, Present, and Future." *Asian Affairs* 53 (3): 711–26.
- Schweig, Meredith. 2022. "Legacy, Agency, and the Voice(s) of Teresa Teng." In *Resounding Taiwan: Musical Reverberations across a Vibrant Island*, edited by Nancy Guy, 213–28.
- Scifo, Salvatore. 2014. "Communication Rights as a Networking Reality: Community Radio in Europe." *Global Transformations in Media and Communication Research*, 164–79.
- Seaver, Nick. 2022. *Computing Taste: Algorithms and the Makers of Music Recommendation*. University of Chicago Press.
- Sharon Kinsella. 1995. "Cuties in Japan." In *Women, Media and Consumption in Japan*, edited by Brian Moeran and Lise Skov, 220–54. University of Hawai'i Press.
- Shen, Puma. 2021. "How China Initiates Information Operations Against Taiwan." *Taiwan Strategists* 12: 19–34.
- Shifu, Wang. 1991. *The Story of the Western Wing*. Edited by Stephen H. West and Wilt L. Idema. Berkeley: University of California Press.
- Silverstone, Roger. 2006. *Media and Morality: On the Rise of the Mediapolis*. Polity.
- Simons, Margaret. 2022. "In Taiwan, Potential Chinese Invasion Leads to Civil Defense Training Boom." *Foreign Policy*, December 22, 2022.
- Singh, Mahesh K. 2024. "Multimedia Application for Forensic Automatic Speaker Recognition from Disguised Voices Using MFCC Feature Extraction and Classification Techniques." *Multimedia Tools and Applications*.
- Singh, Pankaj, and Shubham Verma. 2015. "Network Topologies." *Journal of Computer Science and Engineering* 1 (5): 223–32.

- Smith, David. 2022. "Pelosi's 'Reckless' Taiwan Visit Deepens US-China Rupture – Why Did She Go?" *The Guardian*, August 7, 2022.
- Smuha, Nathalie A. 2021. "From a 'Race to AI' to a 'Race to AI Regulation': Regulatory Competition for Artificial Intelligence." *Law, Innovation and Technology* 13 (1): 57–84.
- Solomon, Thomas. 2012. "Where Is the Postcolonial in Ethnomusicology?" *Ethnomusicology in East Africa: Perspectives from Uganda and Beyond*, 216–51.
- Spivak, Gayatri Chakravorty, and Ellen Rooney. 1989. "In a Word." *Differences* 1 (2): 124–56.
- Staats, Jennifer, Johnny Walsh, and Rosarie Tucci. 2019. "A Primer on Multi-Track Diplomacy: How Does It Work?" United States Institute of Peace.
- Steger, Manfred B. 2009. *The Rise of the Global Imaginary: Political Ideologies from the French Revolution to the Global War on Terror*. Oxford University Press.
- Sterne, Jonathan. 2003. *The Audible Past: Cultural Origins of Sound Reproduction*. Duke University Press.
- . 2012. *MP3: The Meaning of a Format*. Duke University Press.
- Sterne, Jonathan, and Mehak Sawhney. 2022. "The Acousmatic Question and the Will to Datafy: Otter.AI, Low-Resource Languages, and the Politics of Machine Listening." *Kalfou: A Journal of Comparative and Relational Ethnic Studies* 8 (1–2): 288–306.
- Stokes, Jacob. 2023. "Resisting China's Gray Zone Military Pressure on Taiwan." *Center for a New American Security*.
- Sunstein, Cass R. 2001. *Republic.Com*. Princeton University Press.
- . 2017. *#Republic: Divided Democracy in the Age of Social Media*. Princeton University Press.
- Szablewicz, Marcella. 2020. *Mapping Digital Game Culture in China: From Internet Addicts to Sports Athletes*. New York: Palgrave Macmillan.
- Szonyi, Michael. 2008. *Cold War Island: Quemoy on the Front Line*. Cambridge University Press.
- . 2013. "Shifting Frontiers: Cross-Strait Relations in the Context of Local Society." In *Mobile Horizons: Dynamics Across the Taiwan Strait*, 74–98. Institute of East Asian Studies Publications, University of California, Berkeley.
- Taeihagh, Araz. 2021. "Governance of Artificial Intelligence." *Policy and Society* 40 (2): 137–57.
- Taylor, Charles. 2004. *Modern Social Imaginaries*. Duke University Press.
- Terren, Ludovic, and Rosa Borge. 2021. "Echo Chambers on Social Media: A Systematic Review of the Literature." *Review of Communication Research* 9 (May): 1–39.
- The Economist*. 2021. "The Most Dangerous Place on Earth," May 2021.
- Thierer, Adam D. 2014. *Permissionless Innovation: The Continuing Case for Comprehensive*

Technological Freedom. George Mason University: Mercatus Center.

- Tobin, Meaghan, Alexandra Stevenson, and Tripp Mickle. 2024. “Has China Lost Its Taste for the iPhone?” *The New York Times*, March 25, 2024.
- Tommassel, Antonela, Juan Manuel Rodriguez, and Daniela Godoy. 2021. “I Want to Break Free! Recommending Friends from Outside the Echo Chamber.” In *RecSys 2021 - 15th ACM Conference on Recommender Systems*, 23–33.
- Tu, Ethan. 2024. “Taiwan AI Labs Releases World’s First GenAI Report on Coordinated Operations in Democratic Nations Exposing Information Manipulators’ Tactic.” Taiwan AI Labs.
- Tung, Wang. 1989. *Banana Paradise (香蕉天堂)*. Central Motion Pictures.
- Verma, Prateek, and Julius O. Smith. 2018. “Neural Style Transfer for Audio Spectrograms.” In *31st Conference on Neural Information Processing Systems (NIPS 2017)*.
- vTaiwan. 2023. “vTaiwan: About.”
- Wang, Jing Adel. 2016. “Affective Listening as a Mode of Coexistence: The Case of China’s Sound Practice.” *Representations* 136 (1): 112–31.
- Wang, Victoria, and John V. Tucker. 2016. “Phatic Systems in Digital Society.” *Technology in Society* 46 (August): 140–48.
- Wang, Yu. 2019. “Listening to the State: Radio and the Technopolitics of Sound in Mao’s China.” University of Toronto.
- Warner, Michael. 2002. “Publics and Counterpublics.” *Public Culture* 14 (1): 49–90.
- Watson, Jini Kim. 2021. *Cold War Reckonings: Authoritarianism and the Genres of Decolonization*. Fordham University Press.
- Webster, Graham. 2019. “A Brief History of the Chinese Internet.” *Logic*, no. 7.
- Wei, Clarissa. 2022. “The Topsy-Turvy End of Zero COVID in Taiwan.” *The New Yorker*, May 2022.
- Weston, Timothy B. 2013. “Taiwanese Newspapers and Politics in China’s Shadow.” In *Mobile Horizons: Dynamics across the Taiwan Strait*, edited by Wen-Hsin Yeh. Institute of East Asian Studies Publications, University of California, Berkeley.
- Wilde, Tyler. 2017. “Thanks to PUBG, China May Now Have More Steam Users than Any Other Country.” *PC Gamer*, December 11, 2017.
- Williams, Dmitri, Scott Caplan, and Li Xiong. 2007a. “Can You Hear Me Now? The Impact of Voice in an Online Gaming Community.” *Human Communication Research* 33 (4): 427–49.
- . 2007b. “Can You Hear Me Now? The Impact of Voice in an Online Gaming Community.” *Human Communication Research* 33 (4): 427–49.
- Wu, Chee-Hann. 2023. “Taiwanese Horror Games and the Ghosts from the Past.” *Taiwan Insight*.

- Wu, Hang. 2021. "Broadcasting Infrastructures and Electromagnetic Fatality: Listening to Enemy Radio in Socialist China." In *Sound Communities in the Asia Pacific*, edited by Lonán Ó Briain; and Min Yen Ong. Bloomsbury.
- Wyse, Lonce. 2017. "Audio Spectrogram Representations for Processing with Convolutional Neural Networks." In *Proceedings of the First International Workshop on Deep Learning and Music Joint with IJCNN*. Anchorage.
- Yang, Guobin. 2009. *The Power of the Internet in China: Citizen Activism Online*. Columbia University Press.
- . 2018. "Demobilizing the Emotions of Online Activism in China: A Civilizing Process." *International Journal of Communication* 12: 1945–65.
- . 2022. *The Wuhan Lockdown*. Columbia University Press.
- Yang, Hsin Yen. 2023. "Cute Politics!: Articulating the Kawaii Aesthetic, Fandom and Political Participation." *Popular Communication* 21 (2): 85–97.
- Ye, Josh. 2023. "China Drafts Rules for Using Facial Recognition Technology." *Reuters*, August 7, 2023.
- Yeh, Wen-Hsin. 2013. "Introduction: Of War and Peace: History and Perspectives." In *Mobile Horizons: Dynamics Across the Taiwan Strait*, 1–24.
- Yeung, Karen. 2019. "Why Worry about Decision-Making by Machine?" In *Algorithmic Regulation*, edited by Karen Yeung and Martin Lodge, 21–48. Oxford University Press/Oxford.
- Yu, Jon. 2021. "The Great Firewall of China, Explained." *Asianometry*. 2021.
- Yurchak, Alexei. 2005. *Everything Was Forever, Until It Was No More*. Princeton University Press.
- Zhan, Jing Da, and Hock Chuan Chan. 2012. "Government Regulation of Online Game Addiction." *Communications of the Association for Information Systems* 30 (13): 187–98.
- Zhe, Zheng-chong [者郑崇]. 2021. "Cong Qiang Wai Dao Boli Xin Zhongguo Geshou 'xiaohua' yu Huang Mingzhi Gandong 'qiang Nei Ren' [From Outside the Wall, 'Heart of Glass' Chinese Singers 'Xiaohua' and Namewee Moved the 'People inside the Wall' 从墙外到玻璃心 中国歌手"小花"与黄明志感动"墙内人" — 普通话主页]." *Radio Free Asia*, November 16, 2021.
- Zheng, Linlin, Jiakang Li, Meng Sun, Xiongwei Zhang, and Thomas Fang Zheng. 2021. "When Automatic Voice Disguise Meets Automatic Speaker Verification." *IEEE Transactions on Information Forensics and Security* 16: 824–37.
- Zheng, Zhang (張正). 2020. "Director's Weekly Diary: A Song for Shortwave" [總台長週記: 給短波的歌.] *Independent Review (獨立評論)*.