Talking to the Midnight Broadcast: Reviving 1990s City Memories with AI

TONGGE YU, Tsinghua University, China and Massachusetts Institute of Technology, USA
FAN XIANG*, Tsinghua University, China

If you could dial into a midnight broadcast from 30 years ago, what would you say? This project explores the potential of artificial intelligence (AI) to reconstruct historical radio broadcasts, focusing on the cultural importance of preserving intangible heritage through sound. Shenzhen's radio culture in the 1990s, influenced by Hong Kong's openness to pop culture, shaped a generation of listeners. Inspired by the iconic 1990s Shenzhen midnight radio program, *A Break at Service Station*, the project revives its host Sophie, a figure beloved for her soothing voice and empathetic presence, as a digital twin with authentic sounds and genuine memories through AI voice cloning and large language models (LLMs). By integrating AI with city memories, the project reimagines a pivotal cultural moment while bridging generational distance—not through nostalgia, but through curiosity about a time of intimate, localized media. It reveals how AI can revive forgotten voices and reshape public memory in a rapidly evolving city.

CCS Concepts: • **Human-centered computing** → **Natural language interfaces**.

Additional Key Words and Phrases: voice cloning, artificial intelligence, human-computer interaction, radio broadcasting, urban memory, cultural heritage preservation

ACM Reference Format:
Tongge Yu and Fan Xiang. 2025. Talking to the Midnight Broadcast: Reviving 1990s City Memories with AI. *Proc. ACM Comput. Graph. Interact. Tech.* 8, 3, Article 41 (August 2025), 8 pages. https://doi.org/10.1145/3736791

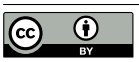
1 Introduction

Sound, unlike visual records, is transient and often overlooked in preserving cultural memory. Shenzhen's cultural evolution [Wang 2016] was influenced by the cross-border exchange of radio and music content from Hong Kong in the late 1980s, a hub for global pop culture and trends. Hong Kong's open-call programs, pop music and dynamic radio formats, helped define Shenzhen's identity as a city balancing rapid modernization with a spirit of community and expression[SCÉRÉN 2006]. In 1990s Shenzhen, a rapidly urbanizing city under China's economic reform, *A Break at Service Station*, a late-night FM89.8 radio program, became a lifeline for young professionals and migrant workers. Hosted by Sophie, a beloved radio host known for her soft voice, playful character, and empathetic presence, the program offered advice, stories, and companionship during a time of profound change.

Today, Sophie's broadcasts survive only as deteriorating cassette tapes, fragmented and filled with static. This project revives her voice and persona using artificial intelligence (AI), showing how emerging technologies can create new ways to experience urban memory and mediated intimacy. We leverage AI voice cloning technology [Ada et al. 2024] and large language models (LLMs)

*Corresponding author.

Authors' Contact Information: Tongge Yu, Tsinghua University, Beijing, China and Massachusetts Institute of Technology, Cambridge, USA, melodyyu6602@gmail.com; Fan Xiang, Academy of Art and Design, Tsinghua University, Beijing, China, xiangfan@tsinghua.edu.cn.



ACM 2577-6193/2025/8-ART41
https://doi.org/10.1145/3736791

[Rothman 2024] to reconstruct Sophie's tonal qualities, conversational style, and personality, ultimately transforming her into an interactive virtual host. The system integrates archival recordings processed through GPT-SoVITS [Jiang et al. 2024], which replicates Sophie's voice with only a few minutes of audio data. To reconstruct her personality and worldview, we use OpenAI's language models [Habib et al. 2024], trained on recorded dialogues and further enriched by Sophie's own writings, including essays and reviews.

Users engage with Sophie through a screen-based interface, designed for mobile devices to simulate the intimate experience of late-night radio calls. By dialing a number, users can talk to Sophie in real time, reconnecting them with the spirit and sounds of 1990s Shenzhen. Fig. 1a illustrates the original radio setting of Sophie in the 1990s, while Fig. 1b captures a participant interacting with her AI-powered counterpart today.

(a) Sophie in a radio station, 1990s Shenzhen

(b) Participant engaging with digital Sophie's voice

Fig. 1. A 30-Year Conversation Across Time

2 Related Work

2.1 Urban Radio and Local Culture

Radio has historically played a crucial role in shaping city memory and fostering cultural identity. Studies such as Sounds of a Migrant City [Lei 2021] explore the transformative role of Shenzhen's radio programming during the 1990s, highlighting its influence on migrant cultural citizenship. Similarly, Talk Radio in Urban China [Xu 1998] examines how talk radio has functioned as a platform for public discourse, reflecting broader societal changes. This work on community radio in Indonesia [Birowo 2011] underscores its potential to empower local culture and democratize access to information.

2.2 AI and Voice Cloning Technologies

The rapid development of AI technologies, particularly in voice cloning, has opened new possibilities for cultural preservation and interaction. Projects such as Amazon's "Grandma Alexa" [Bluestein 2022] and AI-Generated Virtual Instructors [Pataranutaporn et al. 2022] demonstrate how AI-enhanced voice technologies can recreate voices from minimal input and foster new modes of engagement.

2.3 Digital Memory in Cultural Storytelling

Sound and interactivity have emerged as transformative tools in cultural storytelling. The LYDSPOR project [Ryding et al. 2023] employs site-specific sound installations to connect cultural heritage with somatic and affective design. Similarly, Recycled Soundscapes [Franinovic and Visell 2004]

bridges past and present through immersive auditory experiences. Tools like Vojo [MIT Open Documentary Lab 2012], initially designed for constrained communication environments, show how voice-based platforms can amplify marginalized voices through transcription and dissemination.

These works provide the technical and cultural backdrop for understanding Sophie not just as a virtual host, but as a digital persona mediating memory, identity, and presence.

3 Methods and Implementation

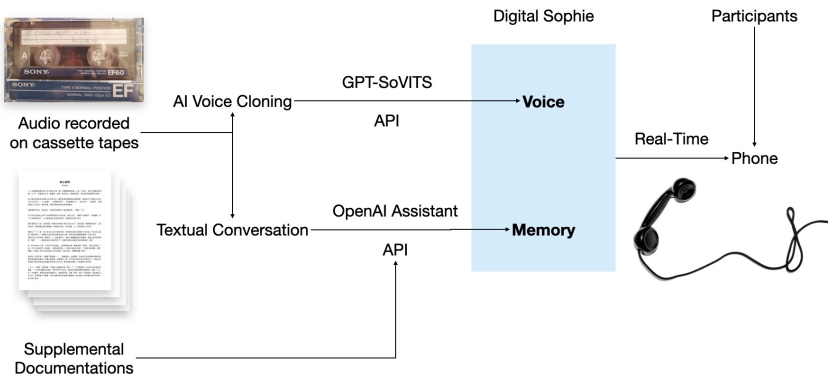


Fig. 2. System Architecture of Digital Sophie

This section introduces the technical system behind Sophie's revival—from the voices and memories that shape her persona to the real-time pipeline that enables her to speak again. We describe how audio archives and personal writings are transformed into conversational behavior, and how users engage with her through a simulated call-in experience.

3.1 Data and Persona

The main dataset is based on 38 episodes of Sophie's original radio broadcasts [Fan 2024], generously provided by Sophie herself. Each episode is approximately one hour long, totaling around 40 hours of audio data recorded between 1993 and 1994. These historical recordings feature themed programs where the host, Sophie, connected with live callers to share personal stories, offer advice, and discuss heartfelt topics.

To enrich her persona, we incorporated nearly 60,000 Chinese characters of Sophie's written works—essays, reviews, and critiques—as supplemental documentation to capture her worldview, narrative voice, and stylistic nuances.

3.2 Voice and Memory

The archival audio was processed along two parallel streams—voice cloning and textual conversion—so that Sophie can both sound and remember like her 1990s self.

Voice cloning. Clean speech segments were used to fine-tune GPT-SoVITS, an open-source model that unifies voice cloning and text-to-speech [Tan 2023]. Only a few minutes of high-quality audio were needed to reproduce Sophie's distinctive timbre, cadence, and emotional nuance with high fidelity [Xue et al. 2024]. At run time, the system calls the GPT-SoVITS API [RVC-Boss 2025] so that Sophie can reply dynamically to user input.

Textual conversion. All episodes were automatically transcribed and indexed to build Sophie's long-term memory. The transcripts feed an embedding database queried through OpenAI's Assistant API [OpenAI 2025], guided by role-specific prompts that emulate her reflective, empathetic storytelling style. This textual memory lets Sophie quote or reflect on past broadcasts, further enhancing the authenticity of the interaction.

3.3 Real-Time Interaction Pipeline

In our system, the interactive experience is designed as a multi-stage pipeline that processes user input, generates responses, and delivers synthesized speech output with minimal perceived latency. The main components of the pipeline are as follows:

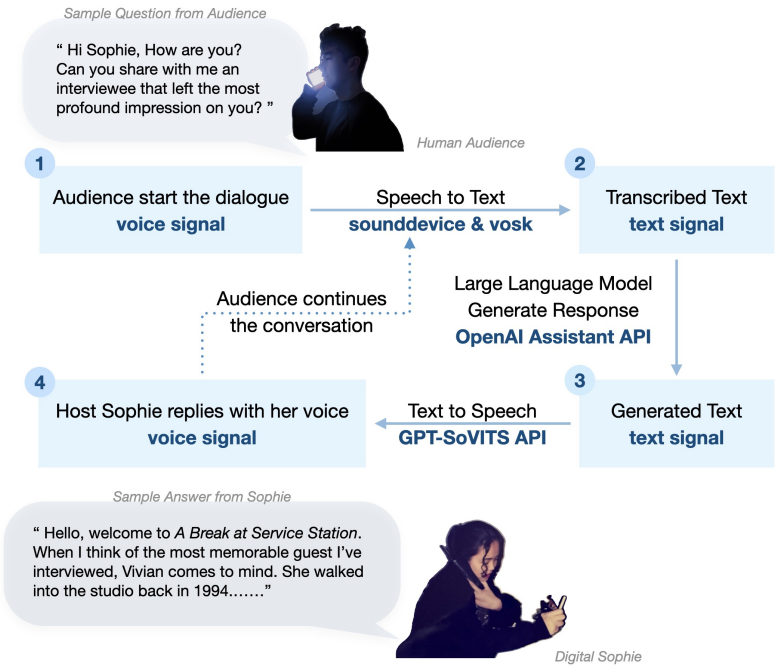


Fig. 3. Real-time dialogue pipeline.

Steps: (1) A spoken question from the audience is captured using `sounddevice`. (2) The audio is transcribed via vosk, an open-source speech-recognition toolkit. (3) The text is processed by the `OpenAI Assistant API`, which generates a response. (4) The response is synthesized into speech using `GPT-SoVITS API` and played back in Sophie's voice.

Speech Recognition: User speech is continuously captured using the Python library `sounddevice` and processed by an offline Vosk model for real-time transcription. We enforce a 2-second minimum recording and apply an RMS-based silence detector before ending capture, ensuring complete utterances even with brief pauses.

Response Generation: The transcribed text is then forwarded to a custom-configured Assistant API through OpenAI's beta threads interface. We synchronously send the transcription to a pre-warmed Assistant API using a lightweight but responsive model such as `gpt-4o-mini`.

Speech Synthesis with Parallel Processing: Once the assistant's text reply is received, it is passed to a dedicated Text-to-Speech (TTS) API for voice generation. For longer responses that may incur delays of over 20 seconds, we split on Chinese punctuation and launch each sentence as an async GPT-SoVITS TTS job. The first segment plays immediately, while subsequent WAVs are queued for overlapping playback.

Latency Optimization: To reduce response latency and improve the user experience, the system adopts the following real-time optimization strategies:

(1) The LLMs generates responses incrementally, enabling partial processing before completion.
(2) Full sentences are detected via terminal punctuation to trigger immediate synthesis.
(3) Using `asyncio`, TTS tasks run asynchronously while playback is handled in a separate thread, allowing synthesis and playback to run in parallel without blocking.

We measured 5–10 seconds from RMS-based end-of-speech detection to first-sentence playback, varying with reply length and complexity.

3.4 Interface and Feedback

Fig. 4. Mobile prototype interface
This screen shows a vintage-style FM dial tuned to 89.8, with a sliding "call" control that users drag to connect. Incoming audio plays through the speaker grill at the bottom.

The physical interactive system of *A Break at Service Station* is designed to provide users with an immersive and tactile way to connect with Sophie's digital twin from the 1990s. This mobile prototype mimics a real phone call: users dial a number to connect and speak with Sophie in real time.

Preliminary testing with participants in Shenzhen revealed diverse emotional and reflective responses. Some described Sophie's voice as warm and comforting, likening it to "the gentlest voice of the night sky". Others reflected on Shenzhen's transformation, recalling its dynamic growth from a city full of opportunities during the 1990s to its current position as a global technology hub. In addition, several participants noted the distinctive communication styles of the era, observing the softer, more thoughtful tone of conversations compared to modern norms. For some, the dialogue prompted reflections on personal histories and family connections, with memories of relatives working in Shenzhen during the city's early days.

4 Discussion and Future Work

User testing revealed that Sophie creates a safe, nostalgic space for open communication. Many participants noted her voice-only presence encouraged intimate sharing akin to 1990s radio call-ins. This empathetic, era-infused interaction model offers applications in therapeutic contexts, from mental-health support to memory-care interventions.

Fig. 5. User workflow interacting with digital Sophie via the mobile radio interface
Steps: (1) The participant opens the app and hears Sophie's broadcast. (2) The default FM dial and "call" slider appear. (3) The participant drags the slider to initiate a live connection. (4) The slider snaps into place, indicating an active call. (5) The participant picks up the handset and asks a question. (6) The participant waits for Sophie's response and continues dialogue.

Looking ahead, we plan to expand beyond individual screens into physical installations—retrofitted phone booths in Shenzhen where citizens can dial a dedicated number and converse with Sophie, providing a tangible window into the city's 30-year transformation. These installations could become interactive urban landmarks, blending nostalgia with technology and fostering a deeper connection between Shenzhen's contemporary residents and their city's history.

5 Conclusion

This project revives *A Break at Service Station* as both a tribute to 1990s Shenzhen and a deeper exploration of how AI can reanimate cultural memory and foster meaningful human connection. By enabling users to engage with Sophie, a recreated radio persona from a transformative era, the project demonstrates how technology can evoke empathy, nostalgia, and reflection while bridging personal and collective memories.

As artists, we reflect on the transient nature of today's digital interactions, from text and voice messages on social media platforms to fleeting exchanges in digital communication apps. These seemingly mundane records of the present may one day become meaningful archives. What

concerns or insights will these records offer to the individuals who revisit them decades from now? Will they, like Sophie's broadcasts, evoke a sense of connection and continuity across time? This project invites us to consider how the ephemeral voices of today might shape the memories and narratives of tomorrow, challenging us to see the everyday as a future archive of cultural and personal significance.

Acknowledgments

We sincerely thank Xiaoqing "Sophie" Xi for providing archival materials and sharing her inspiring stories, and we appreciate all user-testing participants for their valuable feedback. Special thanks go to Xianglin Ji for his on-screen role in the project video, and to Yu Fu for expertly editing portions of the video. Their support has been indispensable to this project.

References

A. A. Ada, S. H. Jørgensen, J. Fritsch, C. A. Le Dantec, A. Vallgårda, L. Jønsson, and S. F. Alaoui. 2024. Cultures of the AI Paralinguistic in Voice Cloning Tools. In *Companion Publication of the 2024 ACM Designing Interactive Systems Conference*. ACM, IT University of Copenhagen, Denmark, 249–252. doi:10.1145/3656156.3663708

Mario Antonius Birowo. 2011. Community radio and the empowerment of local culture in Indonesia. In *Politics and the Media in Twenty-First Century Indonesia*, Sen Hill (Ed.). Routledge, London, UK, 329–345. http://e-journal.uajy.ac.id/id/eprint/28779

Adam Bluestein. 2022. Why Amazon's 'dead grandma' Alexa is just the beginning for voice cloning. https://www.fastcompany.com/90775427/amazon-grandma-alexa-evolution-text-to-speech.

Xiang Fan. 2024. 1994 Shenzhen Midnight: A Break at Service Station. https://www.ximalaya.com/album/80003993. Audio materials uploaded by the author to Ximalaya platform. Accessed: 2025-01-16.

Karmen Franinovic and Yon Visell. 2004. Recycled Soundscapes. In *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques (DIS '04)*. Association for Computing Machinery, New York, NY, USA, 317. doi:10.1145/1013115.1013167

Henry Habib, Sarah McKay, and Peter Siegel. 2024. *OpenAI API Cookbook: Build Intelligent Applications Including Chatbots, Virtual Assistants, and Content Generators* (1st ed.). Packt Publishing, Limited, Birmingham, UK.

Yu Jiang, Tian Wang, Hao Wang, Chen Gong, Qi Liu, Zhi Huang, Li Wang, and Jun Dang. 2024. Expressive Text-to-Speech with Contextual Background for ICAGC 2024. In *2024 IEEE 14th International Symposium on Chinese Spoken Language Processing (ISCSLP)*. IEEE, Beijing, China, 611–615. doi:10.1109/ISCSLP63861.2024.10800495

W. Lei. 2021. Sounds of a migrant city: radio representation and cultural citizenship in the case of Shenzhen, China. *Citizenship Studies* 25, 8 (2021), 1042–1057. doi:10.1080/13621025.2021.1981826

MIT Open Documentary Lab. 2012. Vojo.co: A Tool for Amplifying Marginalized Voices. https://docubase.mit.edu/tools/vojo-co/. Accessed: 2025-01-16.

OpenAI. 2025. OpenAI API Reference: Assistants. https://platform.openai.com/docs/api-reference/assistants. Accessed: 2025-01-16.

P. Pataranutaporn, J. Leong, V. Danry, A. P. Lawson, P. Maes, and M. Sra. 2022. AI-Generated Virtual Instructors Based on Liked or Admired People Can Improve Motivation and Foster Positive Emotions for Learning. In *2022 IEEE Frontiers in Education Conference (FIE)*. IEEE, Uppsala, Sweden, 1–9. doi:10.1109/FIE56618.2022.9962478

Denis Rothman. 2024. *Transformers for Natural Language Processing and Computer Vision: Explore Generative AI and Large Language Models with Hugging Face, ChatGPT, GPT-4V, and DALL-E 3* (3rd ed.). Packt Publishing, Birmingham, UK.

RVC-Boss. 2025. GPT-SoVITS: api.py. https://github.com/RVC-Boss/GPT-SoVITS/blob/main/api.py. Accessed: 2025-01-16.

Karin Ryding, Vasiliki Tsaknaki, Stina Marie Hasse Jørgensen, and Jonas Fritsch. 2023. LYDSPOR: An Urban Sound Experience Weaving Together Past and Present Through Vibrating Bodies. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 184, 16 pages. doi:10.1145/3544548.3581523

SCÉRÉN. 2006. Hong Kong–Shenzhen: The Little Wall of China. Video, New York, NY: Filmakers Library.

Xu Tan. 2023. *Neural Text-to-Speech Synthesis* (1st ed.). Springer Nature Singapore, Singapore. doi:10.1007/978-981-99-0827-1

Da Wei David Wang. 2016. *Urban Villages in the New China: Case of Shenzhen* (1st ed.). Palgrave Macmillan US, New York, NY, USA. doi:10.1057/978-1-137-50426-5

Hua Xu. 1998. Talk Radio in Urban China: Implications for the Public Sphere. In *Communication and Culture: China and the World Entering the 21st Century*, D. Ray Heisey and Wenxiang Gong (Eds.). Brill, Leiden, Netherlands, 329–345. doi:10.1163/9789004455023_021

Jinlong Xue, Yayue Deng, Yingming Gao, and Ya Li. 2024. Retrieval Augmented Generation in Prompt-based Text-to-Speech Synthesis with Context-Aware Contrastive Language-Audio Pretraining. https://arxiv.org/abs/2406.03714. arXiv:2406.03714 Interspeech 2024 (accepted)..