Vocal Organ

Apparatus for Generating Voice and Images

Ziyi Li

Digital Media (M.A.)

University of the Arts Bremen

Prof. Dr. Andrea Sick

Prof. Ralf Baecker and Prof. Natascha Sadr Haghighian

January 15, 2026

Contents

Introduction

The interest in this project initially stemmed from my reflections on my mother: being congenitally deaf, how did she learn to vocalize, master language, and communicate with others? Regarding this atypical cognitive mode, I was curious whether something could be learned and discovered from it: in the absence of hearing, how does one use existing senses to reorganize the mode of perception? How can alternative principles be found to activate the articulators for speech? Why are humans able to naturally acquire speech through hearing? Why do such significant differences exist in pronunciation? Do articulation and action occur simultaneously? These puzzles and complex emotions led me to initiate this artistic research.

I envisioned whether, through this artistic research, it is possible to explore those latent bodily perceptual abilities that are overlooked by people. For instance, most people tend to depend heavily on their sense of hearing, whereas my mother's experience reveals how touch, vision, and muscle memory can substitute for hearing. At the same time, I attempt to examine the human voice from the perspective of materiality: the human voice is not merely sound waves being heard, but the precise collaboration of bodily movement and muscles; that is, the human voice is not just sound, but a deliberate bodily action determined and influenced by a material basis. Furthermore, does my mother's experience reveal an "alternative cognition," thereby challenging our definition of communication? I am curious if a cognitive framework can be discovered from this, like a unique logic revealing how consciousness constructs connections in silence, or the limits and possibilities of human adaptive perception.

My microscopic inquiry into bodily perceptual potential and cognitive reorganization did not stop at a single individual case. In fact, the process of my mother reconstructing perception on a physiological level constitutes a profound intertextuality with the sense of urgency I feel in seeking methodological reconstruction on an artistic level. To adapt to the world, my mother had to find new logic and tools within her lack; similarly, facing current social processes, art also faces a failure of perception and a crisis of paradigms. This projection of perspective from the "microscopic body" to the "macroscopic context" prompts me to re-examine the function and position of the artistic ontology.

The theoretical dimension of this article originates precisely from my reflections on the changes and contrasts of art in current social processes compared to the past, and also from inquiries into my own work: Where is my working method positioned? Where is art positioned? And where will it be in the future?

Regarding the current direction of contemporary art, the seemingly diverse discussion themes and conceptual forms are actually becoming increasingly homogenized. This makes it increasingly limited and constitutes a structure completely different from Modernist art. Pursuing new reorganization under a deconstructionism that counters nihilism often traps us deeper in bewilderment and paradoxes; coupled with the iteration of technology and media, art is challenged in both material and conceptual terms. The original foundations of Western art are being constantly dismantled and critiqued amidst the game between economic markets and national politics. Every aspect indicates that the field of art needs new methods to adapt to the present and face the future. In fact, we have been in this stage for a long time.

There exists a core issue here: although contemporary art has expanded into other disciplinary fields, it often lacks genuine adherence to their internal logic and depth. By comparison, art during the Modernist period did not attempt to discuss too many external

matters, but focused on the autonomy of art forms and media themselves. However, this extensive but sometimes superficial expansion of contemporary art has brought much trouble as it is not only difficult to be recognized by external fields but also causes disorder in internal artistic standards. Therefore, we need effective methods and tools to grant legitimacy to this "expansion," making it both acceptable externally and inheritable internally.

Based on my personal way of creation and thinking, I adopted the following path: regarding myself as a source point for collecting signals from the world, and as an individual with a comprehensive perceptual perspective, to dissect current practices and theories. From this observation point, the first signal I captured is the intense drive of media and scientific progressiveness on artistic transformation. This is filled with complex tensions between the artistic aura, human centricity, and external technology. It is precisely within this shifting scenario and material influence triggered by technological iteration that old definitions have failed. We are forced to ask: facing a reality that changes in real-time, how do we understand and define art in a real time?

Based on this, through practice and cognition of the world, I offer an understanding of art: can art be understood as a methodology that changes in real-time with the situation? It is no longer fixed to a rigid structure or a single historical background, but is a dynamic balance. This avoids attempting to maintain a generalized equality with a generic tone of "diverse but mutually isolated," which thereby causes invisible violence and imbalance.

Chapter 1　My Artistic Practices and Research on Voice

This chapter articulates and analyzes phenomena and questions that surfaced during the investigation by presenting three distinct directions of my personal artistic practice revolving around the voice, serving as a foundation for subsequent reflection and argumentation.

1.1 Project Ⅰ: The Body and the Instrument

1.1.1 The Physics of Voice: From Nature to Structure

The first artistic practice originated from my contemplation on the mechanics of vocalization: How is sound produced? From the outset, this inquiry is situated within a dual context: the perspective of anthropocentrism and the relationship between humans and their environment.

In physics, sound is a vibration that requires a medium for propagation.[1] The emergence of sound is related to Earth's formation and evolution. The medium for sound propagation was prepared once the atmosphere and oceans formed. Before the appearance of life, Earth was a world filled with primitive natural sounds such as wind from flowing air, waves or rain from water's movement and impact, and thunder from lightning instantaneously heating and rapidly expanding air.[2] All these are sounds generated by vibration.

[1] David Halliday, Robert Resnick, and Jearl Walker, "Fundamentals of Physics", 10th ed. (Hoboken, NJ: John Wiley & Sons, 2013).
[2] Bernie Krause, "The Great Animal Orchestra: Finding the Origins of Music in the World's Wild Places" (New York: Little, Brown and Company, 2012).

However, these vibrations, originally limited to the physical world, underwent a significant transformation with the emergence of life. As life evolved, the Earth's soundscape became increasingly complex.[3] Sound was no longer just a byproduct of atmospheric and geological movements, but began to serve the vital function of biological communication.[4]

When humans entered this world as listeners, sound ceased to be a purely physical phenomenon and became an object of perception. For early humans, it functioned not only as a signal for survival but also as a medium for expression. Throughout the long process of evolution, humanity was not content to simply receive nature's vibrations passively. Instead, we began to take control, attempting to convert the unpredictable sounds of the external world into controlled personal expression. This exploration inevitably started with the closest instrument available, our own bodies, before eventually extending to the materials of the outside world.

Naturally, this exploration began with internal experimentation. Before the invention of artifacts, humans utilized their own bodies to create sound and rhythm through clapping, stomping, and vocalizing. It was from this primal basis of self-vocalization that the concept of the musical instrument evolved.[5] The birth of external tools stemmed from humanity's drive to imitate, explore, and expand upon the sounds of nature.

Regarding the origin of these external instruments, a compelling body of hypothesis suggests they evolved serendipitously from the utilitarian tools of hunting and survival. It is hypothesized that early humans might have accidentally thrown a stone with a hollow cavity

[3] Bryan C. Pijanowski et al., "Soundscape Ecology: The Science of Sound in the Landscape", BioScience 61, no. 3 (March 2011): 203–16, https://doi.org/10.1525/bio.2011.61.3.6.
[4] Jack W. Bradbury and Sandra L. Vehrencamp, "Principles of Animal Communication", 2nd ed. (Sunderland, MA: Sinauer Associates, 2011).
[5] Curt Sachs, "The History of Musical Instruments" (New York: W. W. Norton & Company, 1940), 25–27.

while hunting. As the object traversed the air in a parabolic arc, it generated a distinct sound that captured their attention.[6] This auditory discovery likely spurred primitive people to explore the acoustic properties of other natural materials, a process of experimentation that gradually evolved into the sophisticated instruments we know today.

Regardless of the specific historical origins, from a modern perspective, we perceive all sound production as remaining anchored to the human body. This allows for a fundamental division between "internally based vocalization" (centered on the vocal cords) and "externally sourced sound" (controlled by the limbs). This distinction aligns with traditional musicological classifications, such as the Sachs-Hornbostel[7] system, which primarily categorizes all sound production methods into two domains: vocal and instrumental.

Following the internal (vocal cord) and external (limb controlled) division, one category of instruments has a special definition because it utilizes both mechanisms. These are the wind instruments, known in music classification as Aerophones, such as flutes, trumpets, saxophones, organ pipes, suonas, and harmonicas (fig. 1.1). The sound production principle for these instruments is that the player uses internal organs (like the lungs and mouth) to generate pressurized gas and directs it into an external instrument structure which then produces musical sound. This mode is therefore different from pure singing and distinct from typical stringed or percussion instruments, constituting an intermediate form between the two.

[6] Jin Jie, "Chinese Music" (Cambridge: Cambridge University Press, 2011), 16–17. This hypothesis specifically traces the origins of globular flutes (such as the Xun) to the Neolithic hunting tool known as the "stone meteor."
[7] Erich M. von Hornbostel and Curt Sachs, "Systematik der Musikinstrumente: Ein Versuch", Zeitschrift für Ethnologie 46, no. 4–5 (1914): 553–90.

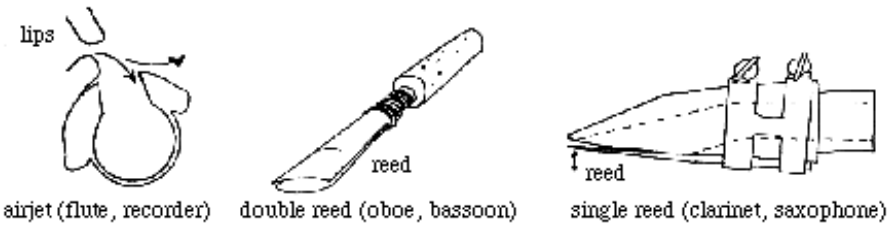

Figure. 1.1. Excitation by air jet or reed. Joe Wolfe, *How do woodwind instruments work?*, UNSW Music Acoustics, accessed January 02, 2026, https://www.phys.unsw.edu.au/jw/woodwind.html.

Aerophones share similarities with the human body in structure and working principles. Some instrument structures highlight this feature. In my investigation the organ pipe is a significant example. As shown in the picture (fig. 1.2), the organ pipe consists of two parts: the foot and the resonator. The foot is located at the base of the pipe, is usually conical, and has a flue at its base where air enters. The resonator supports the air oscillation generated at the mouth, which is the horizontal opening where the resonator connects to the foot. The voicing, resonator length, and volume collectively determine the pipe's fundamental pitch. The conical tapering of the body determines the overblown pitch.

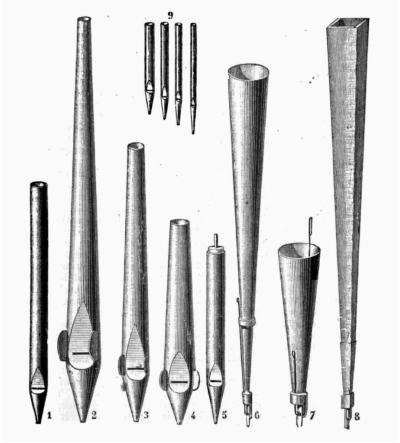

Figure 1.2. Organ pipe shapes. William Henry Stone, "Elementary Lessons on Sound" (London: Macmillan & Co., 1879), 165, fig. 56.

Between the foot and the resonator (fig. 1.3), the side of the pipe body (including the mouth) is flat. A metal or wooden plate called a "languid" is fixed horizontally here, blocking the airway, leaving only a "flue" slit next to the mouth. This design forces the air

into a thin sheet stream, directing it precisely at the mouth. Thin metal or wooden plates called "ears" can be attached to the sides of the mouth for tuning. A horizontal round bar called a "roller" or "beard" may also be fixed to the pipe body to ensure prompt speech.

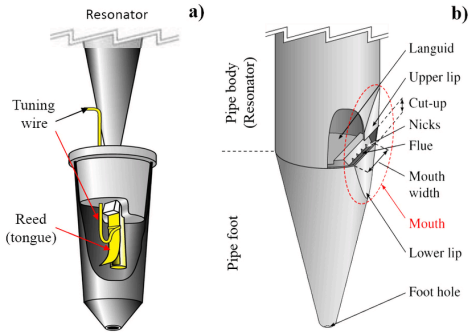

Figure 1.3. The parts of a reed (lingual; a) and a flue (labial; b) organ pipe. Judit Angster, Péter Rucz, and András Miklós, *Acoustics of Organ Pipes and Future Trends in the Research*, Acoustics Today 13, no. 1 (Spring 2017): 11.

The organ pipe shows us an adjustable hybrid organ that integrates human articulatory structures and auditory principles. Many of its structures are analogous to human sensory organs and are directly named after body parts. In my view, the organ pipe is not a one to one copy of human organs but rather an abstract understanding of body organs combined with physical phenomena. It creates a structure that can produce different pitches and timbres through regulation.

1.1.2 Peruvian Whistling Vessels and Me

I became interested in this abstract reunderstanding of the body and the attempt to reproduce (vocalization) through other means. I thus began to pay more attention to movements related to sound's origin like breathing, vibration, and resonance. During this

time I discovered an ancient instrument from Peru known as Peruvian Whistling Vessels[8] (fig. 1.4-6). They are pottery vessels that can produce sound. They vary in shape and size and can emit different sounds. Their sound production principle depends on their construction and the complex hydraulic engineering techniques used inside the container (fig. 1.7). The Peruvian Whistling Vessels have an internal sound structure based on a hydraulic whistle. Its working principle is that as water flows through the vessel it pushes air through a whistle hole, activating the whistle mechanism to produce sound. Their sounds often seem to imitate animal roars and calls.

Figure. 1.4-6. Double-Whistle Vessel with Sea Birds. Sicán (Lambayeque) culture, northern Peru, 900–1100, earthenware, The Walters Art Museum, Baltimore (Anonymous gift, 2009, acc. no. 48.2826). Ellen Hoobler and Sonia Matheus, *The Art of Knowing When Not to Wet Your Whistle: Multidisciplinary Research into the Use of a Sicán Whistling Vessel in the Museum Context*, The Walters Journal 75 (2017): 12, fig. 1.

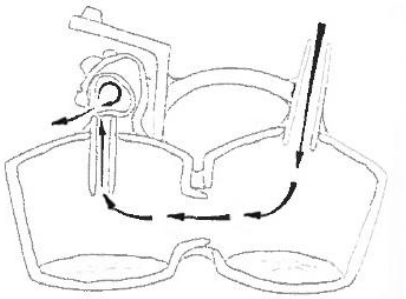

Figure. 1.7. Diagram of a double-chambered whistling vessel. Hoobler and Matheus, *The Art of Knowing When Not to Wet Your Whistle*, fig. 5.

[8] Daniel K. Statnekov, "Animated Earth: A Story of Peruvian Whistles and Transformation", 2nd ed. (Berkeley, CA: North Atlantic Books, 2003), 45.

I also attempted to handcraft several Peruvian Whistling Vessels (fig. 1.8), but the results were disappointing—they sounded like someone snoring. The sound was very muddy and not clear. I suspect this was due to an incomplete water seal and the whistle's aperture not being sharp or smooth enough. However, this unexpected result gave me some new ideas.

Figure. 1.8. *Breathing*. Clay, water, three sets, 25x8x16cm. ©Ziyi Li, 2021.

Before I started making these vessels, I did not consult any texts or structural diagrams regarding their mechanism. I relied solely on an online video with an incomplete demonstration that focused only on the exterior shape. To some extent, this made the task difficult and perhaps seemed imprudent, but now I must express my gratitude for that impulse, and I will explain what it brought me.

From the various online videos I watched, I could see that water is poured in, and the vessel only produces sound when tilted at a specific angle. Combining this with its shape, I deduced that the tubular structure at the bottom serves as a channel connecting two hollow volumes. As water flows from one side to the other, sound is generated only during a unidirectional tilt. Based on this, I reasoned that the upper structures of the two hollow volumes must differ: one side must be open and direct, while the other is a closed structure

for sound. Linking the core movement of "tilting" with the fluidity of water, I immediately associated this with using the change in water volume to compress the hollow volume.

Ultimately, my logic proved correct, although the spatial design and acoustic mechanism of the original vessels were far more complex than I had imagined; I had simply added a structure based on a common whistle. This result was not unsatisfactory; rather, the exploration process fueled my imagination. Inferring internal principles from external form and function offers a different path—a clumsy yet natural method of discovery.

However, from the external appearance of the pottery, I was able to roughly infer its chamber structure and connection method. This creation process parallels my thinking about how my mother vocalizes. She, too, uses her naked eye to observe the external appearance of others speaking to reverse engineer the method of vocalization. The result is just like my Peruvian vessels: the imitation approaches the feeling of the original sound but does not completely replicate it. Instead, it creates a new kind of sound. This discrepancy caused by imitation and replication gives rise to an alternative space and time.

1.1.3 Kempelen, Riches, and their Speaking Machines

In the domain of exploring the replication and imitation of the human voice, German artist Martin Riches presents a deconstruction and reconstitution of the human vocal apparatus through a series of mechanical installations. His seminal work, *The Talking Machine* (fig. 1.9), functions fundamentally as an acoustic speech synthesizer. Adopting the operational mechanics of the pipe organ, the device physically simulates the structural resonance cavities of the human vocal tract. By integrating sound producing components such as reeds or whistles, it achieves speech through the precise coordination of air valves controlled by a computer.

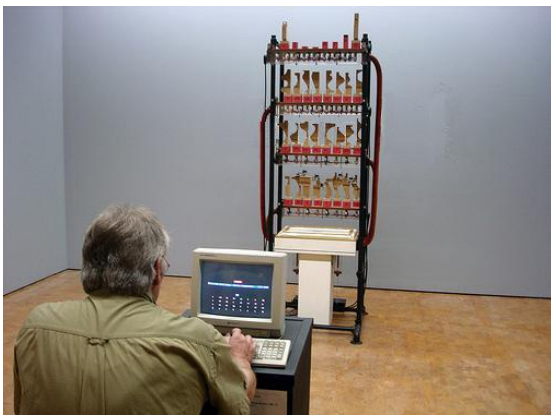

Figure 1.9. The Talking Machine. ©Martin Riches, accessed October 11, 2025, https://martinriches.de/.

Regarding his historical influences, Martin Riches acknowledges a profound debt to Wolfgang von Kempelen, the 18th century inventor who constructed his renowned *Speaking Machine* in 1791 (fig. 1.10), yet he explicitly notes that his work adopts an acoustic approach which was categorically discarded by his predecessor: the allocation of a separate pipe for each individual sound.[9] Riches' creation builds upon the accumulated work of numerous predecessors, combining computer control to strive for a perfect simulation of a speaking machine. By assembling individual syllables to form words and integrating remote control, he achieved a state of perfection in this specific domain. However, as Kempelen concluded after long-term experimentation in his book *The Mechanism of Human Speech*[10], and as Riches himself noted, this approach has limitations. In the fifth chapter of his book, Kempelen recounted his process of making a speaking machine and the problems he encountered[11]. The core issue was that synthesizing coherent sentences by mechanically splicing independent sounds through multiple pipes could never achieve a satisfactory, bionic vocalization. This is because human vocalization relies on a single vocal tract, which is one mouth. Kempelen proposed that the essence of speech is airflow formed within a single channel through the continuous, dynamic changes of the mouth's shape; if a separate

[9] Martin Riches, *The Talking Machine*, accessed December 5, 2025, https://martinriches.de/talk.html.
[10] Wolfgang von Kempelen, "Mechanismus der menschlichen Sprache / The Mechanism of Human Speech", ed. Fabian Brackhane, Richard Sproat, and Jürgen Trouvain (Dresden: TUDpress, 2017).
[11] Kempelen, "Mechanismus der menschlichen Sprache", 388-408.

pipe were built for each letter, like a pipe organ, unavoidable pauses or noises would occur between letters, destroying the coherence of speech. Therefore, Riches' speaking machine and Kempelen's machine represent two distinctly different approaches and fundamentally different understandings of the vocal system. I wish to further discuss why Kempelen's research led to a single channel, while Riches persisted with multi-channel research.

Figure. 1.10. Wolfgang von Kempelen's Speaking Machine (Sprechapparat). Presumed date c. 1800. Collection of the ©Deutsches Museum, Munich, accessed January 15, 2026, https://www.deutsches-museum.de/museum/sammlung/highlights/sprechapparat.

In his book[12], Kempelen started from the language used in daily human communication. Driven by confusion regarding humans and language itself, he proceeded to analyze the physiological organs that facilitate speech and the methods of articulation, ultimately explaining how to construct this speaking machine mechanically. We might refer to Kempelen's work that predates the Speaking Machine: *The Turk* (the chess-playing automaton) (Figure 1.11). He mentioned that in 1769, while dedicated to making the chess machine, he had already begun researching which musical instruments most closely resembled the human voice. He investigated and studied a series of objects capable of emitting tones or sounds, searching for something with a vocal mechanism similar to the human glottis. He mentioned the "vox humana" stop on the pipe organ, composed of various clarinet reeds large and small, but abandoned it because its imitation of the human voice was

[12] Kempelen, "Mechanismus der menschlichen Sprache".

very poor. It was not until he accidentally discovered the bagpipe (also known as the "pipesack") in the countryside that he found a timbre he considered most like the human voice. I speculate the reason lies in the fluidity of the material: the vibration produced by the smoothness of the leather differs significantly from the strong echoes produced by rigid materials.

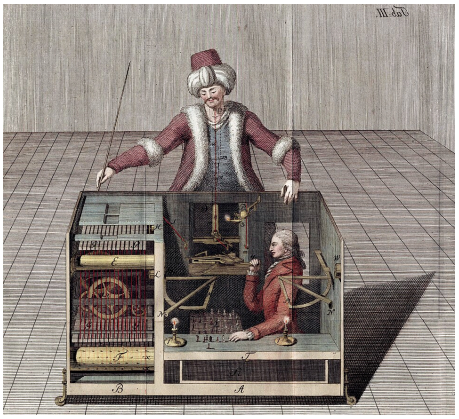

Figure. 1.11. *The Turk* (the chess-playing automaton). Humboldt University Library, accessed December 10, 2025, https://www.digi-hub.de/viewer/image/BV041097321/65/.

Some scholars[13] suggest that the fifth chapter is actually the core of Kempelen's entire speaking machine project, but his lack of extensive description and academic verification has caused trouble for subsequent research. Given my own practice mentioned earlier in this chapter, I, like Kempelen, conducted a series of instrument investigations based on intuitive experience. Combining this with his writing, I boldly hypothesize that the entire process of investigating vocal simulation was filled with accidental occurrences and external interference. This part is difficult to argue academically; it is saturated with personal practice and the circumstances of one's surroundings. For instance, he mentions hitting a bottleneck when trying to replicate the principle of the bagpipe to simulate the human voice, then turning to buy organ pipes from a manufacturer that could control different human-like

[13] Fabian Brackhane, Richard Sproat, and Jürgen Trouvain, *The 'Mechanism of Human Speech': Motivation* in "Mechanismus der menschlichen Sprache" by Wolfgang von Kempelen, ed. Fabian Brackhane, Richard Sproat, and Jürgen Trouvain (Dresden: TUDpress, 2017), CX–CXII.

tones via a keyboard[14] (Fig. 1.12). This experience, combined with my understanding of his entire R&D process, precisely led him in what he later deemed the "wrong" direction—the very method Martin Riches later dedicated himself to. Yet, this "wrong" direction was a crucial validation for bionic voice invention. It was the starting point for distinguishing between multi-channel and single-channel speech generation, and the key turning point that allowed Kempelen's single-channel machine to proceed. Although he did not fully display the internal mechanism of his device to the public, posterity simulated the device based on his methods, proving its feasibility.

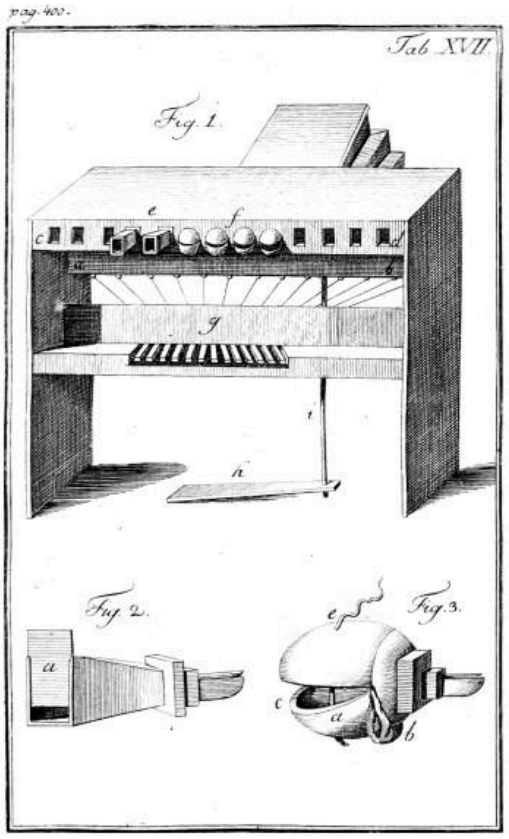

Figure. 1.12. The "human voice" mechanism acquired by Kempelen from an organ builder. The device features a wind chest (a–b) with thirteen wooden mouthpieces. Details d–e and f–g illustrate the reed and shallot mechanism, Wolfgang von Kempelen, "Mechanismus der menschlichen Sprache".

[14] Kempelen, "Mechanismus der menschlichen Sprache".

Returning to Martin Riches' creation, and associating it with Kempelen's historic Turk and Speaking Machine, I am led to a further confusion: Is decomposing different human sounds into different outlets and controlling them any different from how humans control their own voices? When humans design a machine that can simulate human speech, we detach from our own bodies and use external technology, using the body to control external technology to mimic our internal workings. Is this the same event, or a completely different one? Are we truly simulating? Do we understand our internal method of vocalization? Is the speaking machine we produce actually simulating human vocalization?

It seems to be a machine designed based on an imagination of internal workings we do not yet fully comprehend, constructed via external technology and controlled indirectly by the body, to simulate the interior we are curious about and imagine, in an attempt to understand "what are we humans?" and "why are we like this?" However, I speculate that this method of assistance through external operation technology will never be a true one-to-one replication. That is to say, by simulating, we manufacture "something else," not the original target. This is like parents conceiving a new life: we create an infant inside the body without actually knowing how it is made, but the infant is born nonetheless. The infant is a brand-new human, not the parents themselves, nor is the infant "simulated." So, what are we simulating? Does simulation imply a process of creating new things and methods based on existing ones? At the same time, I believe Riches's work intuitively reflects how the general public understood and used computers in the early 21st century, namely, regarding input instructions and output results. This represents the primary, surface-level performance of computers, but it also brings about a misunderstanding of computer capabilities and limits operation. Currently, artificial intelligence based on Large Language Models is changing this situation.

1.2 Project Ⅱ: Hyoid Bone and Articulatory Organs

In this project, I continue to explore the theme: "How exactly does a deaf mother learn to speak?" Through the research and artistic practice of the previous project, I found that objects possessing spatial and postural qualities, such as containers, instruments, and machines kept reappearing. This triggered a reflection on the body: Is the body merely a basic element of physical vibration in phonation? What does the body mean for sound?

Another question arising from this same doubt will also be addressed in this chapter. Distinct from the physical sound phenomena and modes of production noted in the previous chapter: How exactly did the mother mobilize the movement of her articulatory organs? In the absence of hearing, what physiological methods did she use to reorganize and mobilize these organs?

Or perhaps, the question should be: Is phonation necessarily linked to hearing? What exactly is the nature of the connection between the senses?

1.2.1 Hyoid Bone

Under my mother's influence, I conducted a series of investigations into deaf education. During speech instruction assisting those with articulation disorders and speechlessness due to hearing impairment, I first learned about the *hyoid bone* (fig. 1.13), a unique bone in the human body. The hyoid bone is also known as the *lingual bone* because its unique horseshoe structure and position can support the tongue root and assist laryngeal

movement, thereby endowing humans with the anatomical basis for producing complex speech and clear articulation.[15]

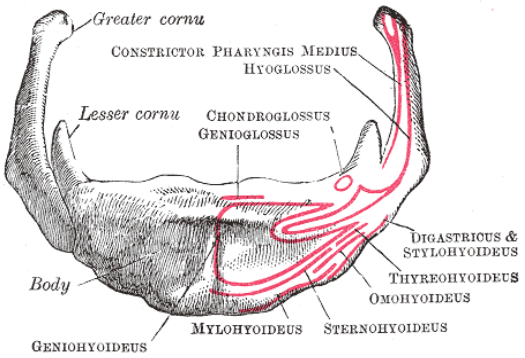

Figure. 1.13. The lesser and greater horns of the hyoid bone. Henry Gray, "Anatomy of the Human Body", 20th ed., ed. Warren H. Lewis (Philadelphia: Lea & Febiger, 1918), Plate 186, accessed November 3, 2025, https://de.wikipedia.org/wiki/Zungenbein#/media/Datei:Gray186.png.

Research indicates that although everyone's hyoid bone is based on the same structural blueprint, it presents unique morphological differences (fig. 1.14). The human hyoid bone consists of three parts; in infancy, they are independent of each other, gradually fusing with growth, though there are special cases where they remain unfused for life.[16] This structural specificity becomes particularly crucial in the cross species comparisons and is often regarded as anatomical evidence for the evolution of a unique human speech system.[17]

[15] Susan Standring, ed., "Gray's Anatomy: The Anatomical Basis of Clinical Practice", 42nd ed. (London: Elsevier, 2020), 582.
[16] Eric Fisher et al., *Hyoid Bone Fusion and Bone Density across the Lifespan: Prediction of Age and Sex*, Forensic Science International 262 (May 2016): 282.
[17] B. Arensburg et al., *A Middle Palaeolithic Human Hyoid Bone,* Nature 338, no. 6218 (April 1989): 758–60, doi:10.1038/338758a0.
.

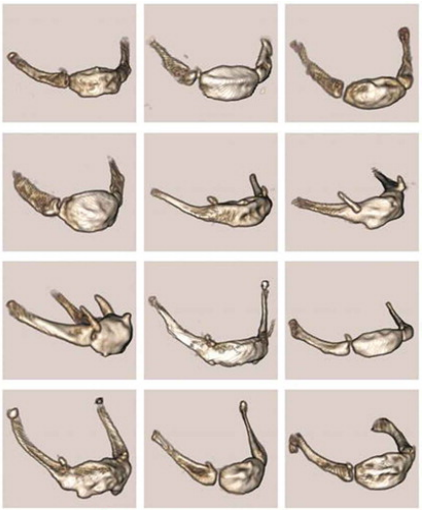

Figure. 1.14. Variations of the hyoid bone. Pettit, Nathan J., and Ronald C. Auvenshine. 2020. *Change of Hyoid Bone Position in Patients Treated for and Resolved of Myofascial Pain*. CRANIO® 38 (2): 74–90. doi:10.1080/08869634.2018.1493178.

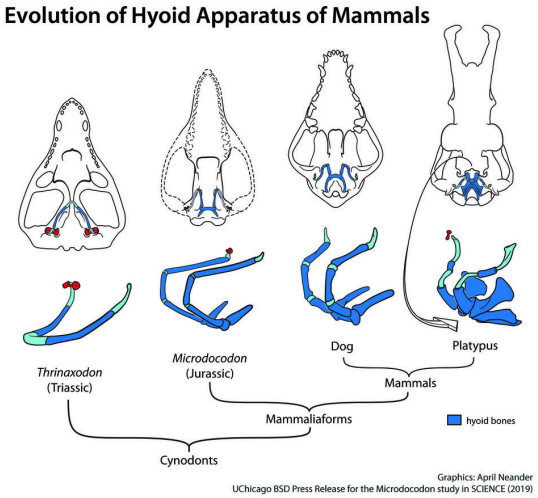

Figure. 1.15. Modern mammals, such as the dog and the platypus, have segmented hyoid bones with mobile joints. The mammalian hyoid's flexibility helps humans and their relatives move their tongues and gulp down chewed food. In stark contrast, reptiles have a rigid, unsegmented, V-shaped hyoid, and often swallow chunks of food whole. April I. Neander, the University of Chicago, Katherine J. Wu, *Mammals' weird way of swallowing is at least 165 million years old*, NOVA, PBS, July 18, 2019, accessed November 3, 2025, https://www.pbs.org/wgbh/nova/article/microdocodon-gracilis/.

For example, the hyoid bone of birds is rooted at the base of the tongue, causing the tongue to be unable to extend out of the mouth; while the horse's hyoid almost runs through the skull, creating a huge resonant cavity and loud sound, but lacking flexibility (fig. 1.15).

In contrast, the evolution of the human hyoid precisely supports flexible tongue movement, thereby giving rise to language and writing.[18] It is precisely this critical anatomical foundation that makes the intricate articulatory system composed of the oral, nasal, and laryngeal cavities possible, turning phonation into a coherent dynamic process that begins with power and is completed through regulation.[19]

The hyoid bone does not articulate with any other bone in the human body; instead, it is suspended from the styloid process of the temporal bone by ligaments and muscles (fig. 1.16). This unique "suspended" structure makes the hyoid bone a core mechanical hub within the neck's soft tissue network.[20] Via the suprahyoid muscles, it connects upward to the mandible and temporal bone, participating in swallowing and jaw movement; via the infrahyoid muscles, it connects downward to the sternum, clavicle, scapula, and thyroid cartilage, assisting in fixing the larynx and aiding respiration.

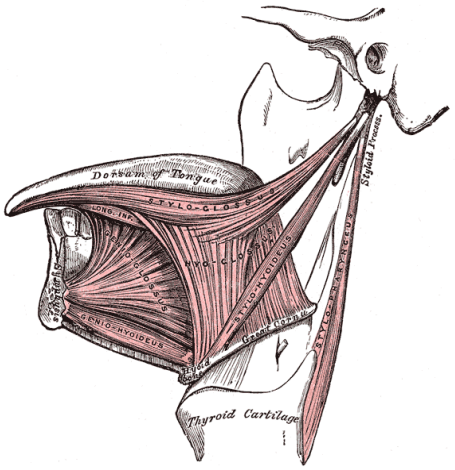

Figure. 1.16. Extrinsic muscles of the tongue, left side. Henry Gray, "Anatomy of the Human Body", 20th ed., ed. Warren H. Lewis (Philadelphia: Lea & Febiger, 1918), Illustration by Henry Vandyke Carter, accessed November 10, 2025, https://commons.wikimedia.org/wiki/File:Gray1019.png.

[18] Philip Lieberman, *The Evolution of Human Speech: Its Anatomical and Neural Bases*, Current Anthropology 48, no. 1 (February 2007): 39–66, doi:10.1086/509092.
[19] Ingo R. Titze, "Principles of Voice Production" (Iowa City: National Center for Voice and Speech, 2000).
[20] Standring, "Gray's Anatomy", 150.

In particular, the omohyoid muscle directly connects the hyoid to the scapula, meaning that tension changes in the shoulder girdle and upper limbs can be transmitted directly to the larynx through myofascial chains.[21] Therefore, the movement trajectory of the hyoid is not limited to local traction in the oral cavity but is the result of the combined action of head posture, cervical curvature, shoulder girdle stability, and respiratory dynamics. Any change in body state involving the aforementioned structures, such as deepened breathing while running, thoracic compression while curling up, or shoulder tension during heavy physical activity will alter the position and stability of the hyoid bone through biomechanical chains, thereby directly intervening in the operational state of the articulatory organs.[22]

In summary, if viewed purely from a local structural perspective, the hyoid is indeed just a bone that helps phonation by regulating the position of the larynx and tongue; however, if viewed from the perspective of whole-body coordination, the hyoid acts more like an intermediary between body movement and verbal activity. Phonation is not an isolated mechanical vibration process in the throat, but an organizational behavior requiring immediate whole body participation. Precisely because the hyoid structurally possesses both the high degree of freedom of "suspension" and tight connections with muscles throughout the body, it can directly transmit the body's overall state (such as the tension or relaxation of posture) to the vocal organs. This special status proves that: phonation is essentially inseparable from the joint participation of the whole body, and the hyoid bone is the hub connecting the bodily background with the act of phonation.

[21] Thomas W. Myers, "Anatomy Trains: Myofascial Meridians for Manual and Movement Therapists", 4th ed. (Edinburgh: Elsevier, 2020).
[22] P. G. C. Kooijman et al., *Comparative Study of the Relation between Laryngeal and Body Posture in Patients with Muscle Tension Dysphonia and Controls,* Journal of Voice 19, no. 3 (September 2005): 366–74.

1.2.2 Under Audism[23], the Body Discipline and Reshaping

From the late 19th to the middle 20th century, Oralism dominated deaf education, advocating the abandonment of sign language in favor of a complete shift to speech acquisition. This period is often regarded as the "Dark Age" of deaf education, marked by the 1880 Milan Conference[24]. The background of this educational orientation deeply reflected the social context of the time, which were widespread social fears regarding assimilation and immigration, the sociological misuse of Darwinian theory, and the strict bodily norms of the Victorian era combined to construct an extreme form of Audism.[25] Shrouded in this ideology and due to the absence of auditory presence, deaf children could not acquire language through the natural path of acoustic imitation and were forced to rely on vision and other sensory modalities to reconstruct language.

During the same period, Alexander Melville Bell created *Visible Speech*.[26] This phonetic notation system, through a set of pictographic symbols simulating the specific movements and positions of oral articulators (such as the tongue, lips, and throat), attempted to guide deaf individuals in pronunciation and accurately transcribe all human languages (fig. 1.17). Although it held epoch-making theoretical significance as the first scientific physiological phonetics scheme, it was ultimately replaced by the International Phonetic Alphabet (IPA) due to its overly abstract and complex symbol design, high writing costs, and lack of compatibility.[27] In my view, the emergence of this system reveals an essence:

---

[23] Tom Humphries, *Communicating across Cultures (Deaf-Hearing) and Language Learning* (PhD diss., Union Graduate School, 1977), 12.
[24] Proceedings of the International Congress on the Education of the Deaf (London: W. H. Allen & Co., 1880), quoted in Harlan Lane, "When the Mind Hears: A History of the Deaf" (New York: Random House, 1984), 393–394.
[25] Douglas C. Baynton, "Forbidden Signs: American Culture and the Campaign against Sign Language" (Chicago: University of Chicago Press, 1996), 15-20.
[26] Alexander Melville Bell, "Visible Speech: The Science of Universal Alphabetics" (London: Simpkin, Marshall & Co., 1867), 15–20.
[27] Margret A. Winzer, "The History of Special Education: From Isolation to Integration" (Washington, DC: Gallaudet University Press, 1993), 192–203.

driven by Audism, the deaf body was alienated into an object requiring "correction" through mechanical training.

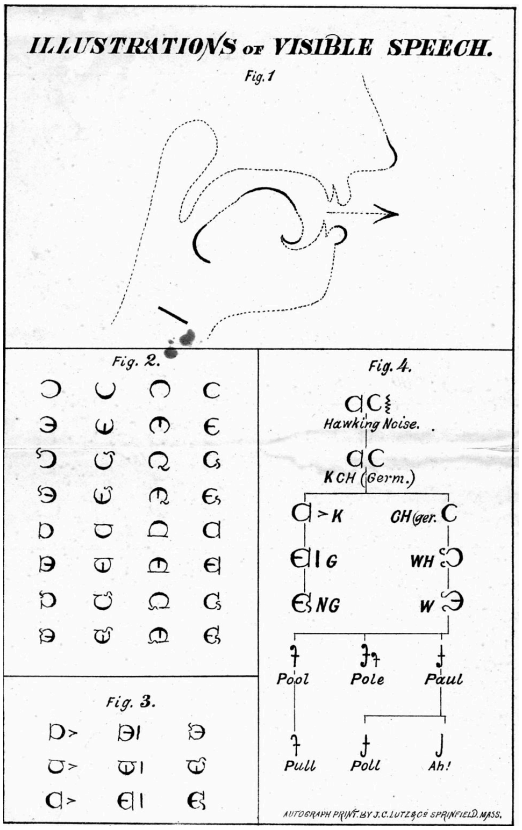

Figure. 1.17. The Fundamental Symbols of Visible Speech. Accessed November 12, 2025, https://commons.wikimedia.org/wiki/File:VisibleSpeech-illustrations.jpg.

To forcibly construct spoken language in a vacuum devoid of auditory feedback, this external social discipline was gradually internalized as an extreme scrutiny of and physical intervention in the articulatory organs. In my research, I noted that the Wooden Tongue Depressor, as a physical medium, was widely used to implement tongue pressure resistance training[28]. The core logic of this method lies in viewing the tongue body as muscle tissue and the hyoid bone as its physiological base; through physical counter pressure, it forces the suprahyoid muscle group to contract and strengthen, thereby pulling the hyoid bone upward, attempting to establish a stable physiological foundation for clear articulation (fig. 1.18).

[28] Caroline A. Yale, "Formation and Development of Elementary English Sounds" (Northampton, MA: Gazette Printing Co., 1892), 8–12.

This is essentially a reverse mechanism of bodily control: through tactile stimulation and resistance at the tip of the tongue, the movement of the hyoid bone is reversed and activated, making the articulatory organs more responsive to neural signals.[29] In this training, I observed a unique path to learning articulation, supported by multisensory perception involving the visual capture of mouth shapes and tactile perception of physical throat vibrations, learning occurs passively through a reverse approach.

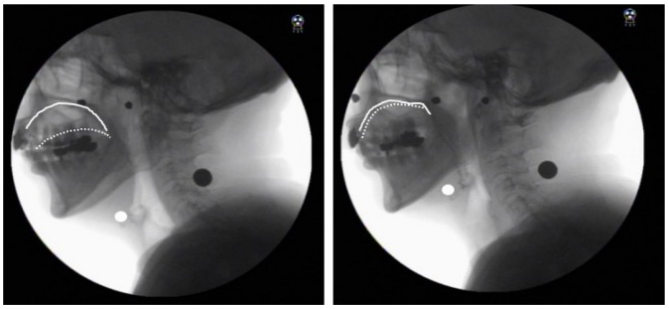

Figure. 1.18. Tongue-pressure resistance training. Notes: The straight line is the palate, the dashed line is the tongue surface, and the white circles represent the hyoid bone. The left panel shows the resting position. The right panel shows the elevated hyoid with the tongue pushing against the palate as hard as possible. Chizuru Namiki et al., *Tongue-Pressure Resistance Training Improves Tongue and Suprahyoid Muscle Functions Simultaneously*, Clinical Interventions in Aging 14 (2019): 601–8, DOI: 10.2147/CIA.S194808.

1.3 Project Ⅲ: The Model which frame the Human Voice

1.3.1 Why a "Model"?

Based on the research and artistic practice of the two aforementioned projects, I propose a core hypothesis: constructing a "model" as a theoretical framework to elucidate the intentions of my art projects and deeply analyze the significance of the Voice.

The discussion begins with the selection of the concept of the "model." We must ask: What is a model? What is its significance? And what inspirations can it offer? In this study,

[29] Chizuru Namiki et al., *Tongue-Pressure Resistance Training Improves Tongue and Suprahyoid Muscle Functions Simultaneously*, Clinical Interventions in Aging 14 (2019): 601–8, DOI: 10.2147/CIA.S194808.

the model is not merely a physical simulation but also a generalizable summary. It possesses applicability and adjustability, and can even be split into multiple sub-models as needed. Combined with artistic creation, this leads to a series of questions: Why does the "model" appear in artistic creation? Is the "model" constructed during the artistic process the same as the "model" used in scientific fields? Why do I place model-making within artistic practice rather than engineering projects? Is this "model" methodology applicable to other forms of artistic creation?

Figure. 1.19. My preliminary concept sketches of the vocal cords for the vocal tract model.

1.3.2 The Basis of the Model and Limitations

To create a resonance channel space that simulates humanoid phonation, I referenced internal oral volume images obtained via MRI to create physical models. My design logic is grounded in the fundamental aerodynamics of phonetics: the energy of the vast majority of languages originates from pulmonary airflow. This airflow is exhaled from the lungs and reaches the first key node—the glottis, where the opening and closing state of the vocal cords determines whether the sound is produced by vibration or simple airflow passage. Subsequently, the raw airflow carrying information enters the upper resonating cavities,

where the soft palate, acting as a "valve," determines its direction: lifting to block the nasal cavity to form oral sounds, or lowering to introduce it into the nasal cavity to form nasal sounds. Finally, the precise quality of the sound depends on the collaboration of various active organs within the oral cavity.

To physically reconstruct this transformation from physiological airflow to meaningful speech, my preliminary plan (Fig. 1.20) was based on medical anatomy, divides the human phonation process into the following functional modules:

- ➔ Power Source: Airflow provided by the lungs.
- ➔ Control Valve: Passing through the larynx and vocal cords, exerting fine control over the volume and state of the airflow.
- ➔ Resonating Cavity: The hollow space constituted by the oral and nasal cavities (Fig. 1.21).
- ➔ Articulator: The tongue, lips, and teeth, responsible for the precise regulation of sound and articulation.

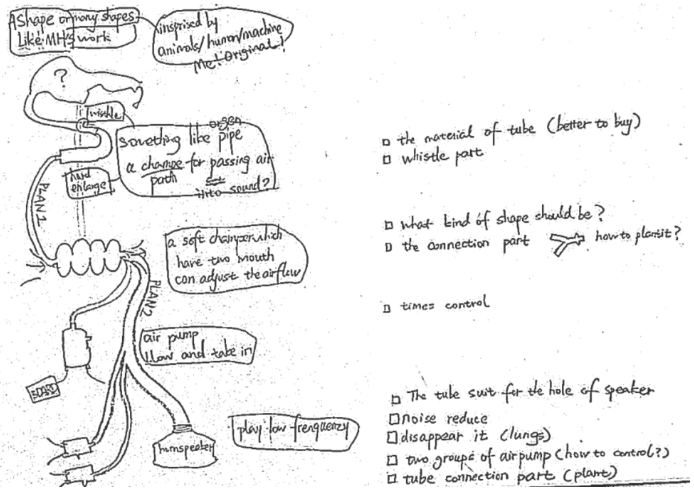

Figure. 1.20. My design process for the simulation of the vocal tract voicing.

Figure. 1.21. My conceptual profile view of a specific movement node within the oral-nasal spaces.

Based on this division, I completed a rudimentary physical simulation in the project: simulating the lungs by using a Raspberry Pi to control an air pump to generate airflow of different cycles; creating sound by modulating airflow using silicone airbags as a variant of vocal cords (Fig. 1.22); and simulating the oral morphology by creating cavities through clay sculpting, mold making, and porcelain slip casting. However, I must reveal the limitations of this division. This regional segmentation of bodily functions harbors inherent cognitive limitations when attempting to deeply understand bodily operations, as the body is an indivisible whole. Therefore, how to understand and confront this "local limitation" is precisely a key component of my work direction, and a point of reflection to be constantly vigilant about during the model's construction process.

Figure. 1.22. I use deformable silicone bladders to mimic the flexible movement of the vocal cords.

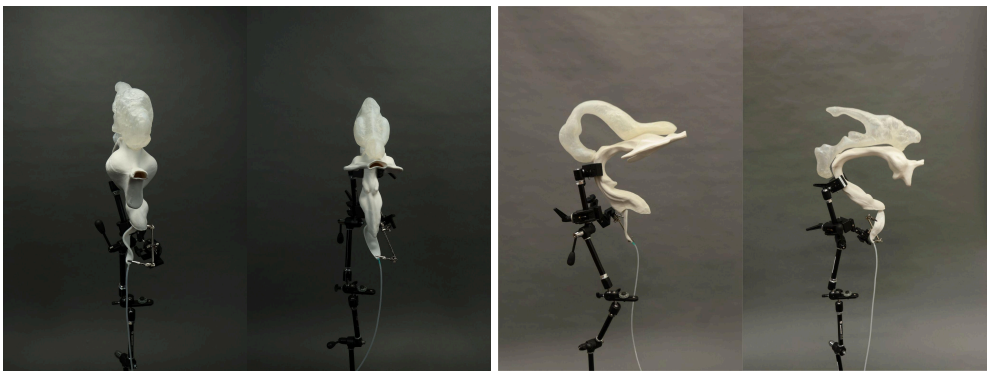

Figure. 1.23-24. *Vocal Tectonics*. Materials: porcelain, soap, raspberry pi, silicone, plastic film, rubber tube, magic arm, platform, metal pipes, mounting brackets, needle, air pump. ©Ziyi Li, 2025.

1.3.2 The Spatial Paradox of Voice

To reshape the invisible negative spaces within oral movement, I observed medical imaging and anatomical diagrams to establish a fabrication logic based on volumetric transformation. This involves a set of spatial relationships that are diametrically opposed yet based on the same volume.

The specific logical sequence of my fabrication and transformation is as follows:

1. First Space (Prototype/Negative Space): This is the actual human oral cavity. It is an invisible, intangible negative space, a void enclosed by the body's internal skeletal and soft tissues.
2. Second Volume (Clay/Positive Entity): To simulate the first space, I created a sculptural volume using clay. This is a materialized replication of the invisible internal oral cavity (Fig. 1.25).
3. Third Volume (Plaster Mold/Negative Entity): I made a plaster mold of the second volume to obtain the third plaster mold. In terms of spatial attributes, this mold itself is equivalent to the skeletal and soft tissues wrapping the oral cavity, which is, the "other side" (container wall) where the first space is located (Fig.1.26-27).
4. Fourth Volume (Porcelain/Final Entity): Using the fluidity of porcelain slip to cast into the third mold, I finally obtained the hollow fourth volume.

This process reveals a paradox of physical existence: the actual volume (porcelain) embodies a space that originally did not exist (the oral void), yet it authentically constructs a physical space. Through the inversion of inner and outer surfaces, I captured that originally untouchable cavity. In this transformation between inner and outer surfaces, actual volumes, and fictional spaces, does a real space actually exist? What clues, witnesses and threads through this process? My answer is the human voice.

Figure. 1.25. One of the models which was created based on various medical image references.

Figure. 1.26-27. The void space revealed within the plaster mold I created for the prototype.

Figure. 1.28. A view from the aperture of the fired ceramic oral cavity.

Chapter 2 The Voice and the Body

Starting from this chapter, I will attempt to construct a perspective based on the mutual verification of theory and practice. By analyzing the findings and reflections derived from my personal artistic practice and research, I aim to elucidate my understanding of the voice and its conceptual extensions.

2.1 Voice, Signifier Chain, and Matter

Saussure redefined the linguistic sign, positing that it "unites, not a thing and a name, but a concept and a sound-image."[30] He argued that a sign comprises two inseparable components, much like the two sides of a single sheet of paper: the signifier and the signified. The signifier is the perceptible component of the sign—that is, its physical form (the "sound-image"). The signified, in turn, is the concept the sign represents. More specifically, this refers to the shared, collective concept understood by a social group.

Whereas Saussure viewed the signifier and the signified as equal and mutually dependent, Lacan, in contrast, posited the primacy of the signifier. Lacan asserted that a fundamental rupture exists between the signifier and the signified; we can never truly capture the ultimate meaning corresponding to any single signifier.

According to Lacan, meaning (the signified) is not permanently attached to a specific signifier. Rather, it is generated within the "signifying chain" through a constant process of slippage, metaphor, and metonymy. One signifier invariably refers to another signifier,

[30] Ferdinand de Saussure, "Course in General Linguistics", ed. Charles Bally and Albert Sechehaye, trans. Wade Baskin (New York: The Philosophical Library, 1959), 66–67.

rather than to a definitive signified. In his *Graph of Desire* (Fig. 2.1), Lacan integrated this structure into psychoanalysis, further concluding that voice serves as the physical substrate of the signifier, yet it does not inherently constitute part of the signifier itself. When the chain of speech is interpreted and a signified is formed by the looping interaction of the two vectors, the voice becomes irrelevant; it is discarded from this process.[31] That is, the instant meaning is received, the voice itself vanishes.

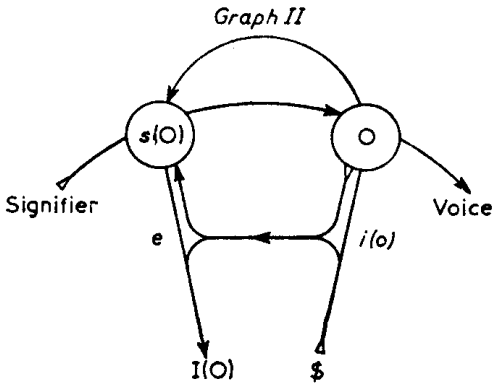

Figure. 2.1. Lacan's Desire Graph. Jacques Lacan, Écrits: The First Complete Edition in English, trans, Bruce Fink, New York: W. W. Norton & Company, 2006, 435.

At this point, the voice is reduced to a pure, residual soundwave, devoid of any signification, which only manifests after the process of comprehension is complete. The voice is thus expelled from the circuit of meaning (Fig. 2.2). This transformation into a residual element, or surplus, is precisely what allows the voice to be constituted as an object at the level of the drive. In this framework, the voice is simultaneously essential and useless. This presents a paradox: we must utilize voice to speak and convey meaning, yet to grasp that meaning, we must discard the voice itself at the very moment of comprehension. It is

[31] John Holland, "The Mark and the Voice", in The Clinic of Transference, ed. Colette Soler (Paris: Éditions du Champ lacanien, 2004), 152-153.

this unique, discarded status that enables the voice, detached from linguistic meaning, to become one of the central objects of human desire, or, more precisely, the drive.

Figure. 2.2. *Freeing the Voice.* Marina Abramović, 1976*,* performance video (Betacam SP, black and white, sound), duration: 28 min. 18 sec. Collection of ZKM | Center for Art and Media Karlsruhe. © VG Bild-Kunst, Bonn 2014, "When you are screaming in this way, without interruption, first you recognize your own voice," Abramovic said of the piece "But later, when you are pushing against your own limits, the voice turns into a sound object.", Dayna Evans, *Pharmakon, Sound Art, and Expressing Bodily Disturbance Through Noise*, Pitchfork, October 23, 2014, accessed November 10, 2025, https://pitchfork.com/thepitch/556-pharmakon-sound-art-and-expressing-bodily-disturbance-through-noise/.

Lacan's Graph of Desire demonstrates that the minimal signifying operation necessarily produces the subject in its retroactive vector, establishing it as an entity that slides along the signifying chain without an exclusive signifier. The subject itself is ungrounded and insubstantial; it is a void necessarily determined by the very nature of the signifier. The voice, therefore, seemingly provides this empty entity with a counterpart, allowing it to gain a certain foothold in positivity—a relationship with presence.[32]

From the perspective of the signifier, materiality is always suspended. The voice here refers not to the content of speech but to the materiality of sound, meaning the voice's timbre, pitch, texture, rhythm, and breath.

[32] Mladen Dolar, "A Voice and Nothing More" (Cambridge, MA: The MIT Press, 2006), 34.

In relation to the signifier chain, the voice plays two key roles. The first role is the material support of the signifier. The signifier chain itself is an abstract structure (like the language system); it requires a material medium to be perceived. The voice is the physical material that carries and conveys the signifier. The second role is the remainder of the signifier chain. This is a more profound connection in Lacanian theory. Lacan regards the voice as a form of Objet petit a. The voice itself is not a signifier. It does not represent anything; it just is. It is the non-signifying part; it is that pure material presence in speech. When you listen to someone speaking, you hear not only the meaning of the words but also his or her voice, that texture that makes one feel anxious, pleased, irritated, or drawn in.

The signifier chain attempts to organize and convey meaning. The voice as the support is responsible for transmitting these signifiers. In this process, the voice itself, as a non-signifying pure presence, simultaneously bypasses meaning and directly touches our desire. The voice detaches from the signifier chain. It reveals the desire or the rift of the real that language and the symbolic order cannot fully conceal. Therefore, the voice is the ghost of the signifier chain. As a material support it is the condition for the signifier chain's occurrence, yet it is also the Other that the signifier chain can never capture.

2.2 Voice, Subject, and Presence

Even the voice, which seemingly conveys meaning most directly, cannot be a *pure present*. The voice is always changing because acoustic waves always experience delay and difference as they propagate through space and time. From this perspective, the aforementioned "material support" is not a natural substance in the traditional sense but rather a psychoanalytic concept of materiality. Here, the word materiality is used to oppose the Symbolic order—that is, the abstraction of language, structure, and meaning.

From the above, we can see that all discussions surrounding the voice are inseparable from its origin, the human subject. In all questions concerning the subject, the voice consistently revolves around the "meaning," "thought," and "soul" it carries, and it is bound to the human body. Based on the discussion thus far, the voice is both material and non-physical; it belongs to the body but is not equivalent to it biologically; it is the carrier of desire but not equivalent to meaning.

To explore the subjectivity of the voice, I will start from the pure physical properties of the voice that we can physically experience and extend the concept of the subject in two different directions.One direction is technological, the external direction. This is the separation of the voice from the body after the invention of the phonograph. Technology can directly record the sound itself, focusing not just on the meaning it carries but on its physical vibration. The voice no longer belongs only to humans. It has become a signal that can be captured, stored, replicated, and manipulated by technical media. The other direction is physiological, the internal direction. This is the voice as sound emitted from the internal structure of the biological body, through air pressure and movement, surrounding the body. Air enters the lungs and produces the voice.

2.3 Voice and Medium

The goal of philosophy is conceptualization, the search for eternal truths capable of infinite repetition. Truth is a non-material, non-worldly idea. To the philosopher, the voice precisely embodies the firsthand transmission of this idea—a pure communication unmediated by any medium. Philosophers therefore believed the voice could avoid the absence of thought.

The example of the ancient Greek philosopher lecturing behind a curtain corroborates this point. He attempted to obscure his physical form with the curtain, intending for the listeners to focus more purely on his voice (thereby reinforcing his intended meaning) rather than being misled by vision. This curtain separated the philosopher's mind and body. The curtain rendered the voice acousmatic, thereby embodying the lecturer's omniscient, omnipresent disciplinary power. The body's absence extended consciousness while the voice brought the illusion of presence.

To compensate for absence, the medium was born. Positioned between presence and absence and due to *Différance*, the medium regarded by philosophers as an obstacle brought forth new bodily experiences, delivering situational experiences based on human senses yet beyond the human body.

Chapter 3 Technologies through the Voice

In this chapter, I will trace the technical history related to the Voice. My aim is to reveal how technology has stepped in, stripped away, reshaped, and redefined the "human voice" by interweaving seemingly unrelated fields. This historical survey is not a mere listing of techniques, but an attempt to understand what kind of ontological shift the human voice has undergone amidst these medial transformations, and whether this brings us new research dimensions.

3.1 Text as the Archive of Silence

In ancient times before the birth of communication technology, writing was the oldest medium for voice transmission. In the Western philosophical tradition, Plato, through the mouth of Socrates in *Phaedrus*, expressed a depreciation of written words, believing that Speech is the direct presence of the soul, while writing is merely its dead shadow.[33] This ideology of "Phonocentrism" was famously deconstructed by contemporary philosopher Jacques Derrida in *Of Grammatology*. [34]

However, shifting the perspective to Media Ontology, writing is precisely the first technical resistance against the fleeting nature of the voice.[35] As a record of the language formed by the voice, although writing cannot reproduce the speaker's immediate Timbre and

[33] Plato, "Phaedrus", trans. Alexander Nehamas and Paul Woodruff (Indianapolis: Hackett Publishing, 1995), 274c–75b.
[34] Jacques Derrida, "Of Grammatology", trans. Gayatri Chakravorty Spivak (Baltimore: Johns Hopkins University Press, 1976; orig. pub. 1967), 11–12.
[35] Walter J. Ong, "Orality and Literacy: The Technologizing of the Word" (London: Routledge, 2002; orig. pub. 1982), 31–32. Ong argues that sound is essentially evanescent, noting that "sound exists only when it is going out of existence."

Prosody, it preserves meaning through symbolic recording.[36] In this media transformation from sound waves to visual symbols, the loss of sensory qualities is inevitable. But writing constructs a new reality: even for forged or tampered texts, their material existence as counterfeits is itself real and verifiable. Writing solidifies flowing sound into a static material carrier available for repeated scrutiny.

3.2 Reconstructions of Old Chinese

Continuing the reflection from the previous section, written records have their limitations, namely "possessing meaning but lacking sound", which subtly catalyzed the special fields of Historical Phonology and "Archaic Articulation." The physical properties of sound determined its non-preservability prior to the invention of the phonograph. Facing this "auditory vacuum," Sinologists represented by Bernhard Karlgren established a rigorous logical system using the rhyme categories of the *Classic of Poetry* (Shijing), the phonetic components of phono-semantic characters, and dialect comparison methods.[37]

This is a reverse method, attempting to reconstruct historically lost sounds within silent texts through logical deduction. The logic here is not simple scientific falsification, but a deduction based on linguistic traces. Researchers attempt to reverse-engineer a theoretical truth within the irreversible unidirectionality of time flow, combining memories of bodily habits such as orally transmitted dialects with the carriers of historical texts. Although this reconstruction pursues 100% authenticity academically, it is essentially a theoretical model built upon modern cognitive frameworks. It is not a physical restoration of

[36] Friedrich A. Kittler, "Gramophone, Film, Typewriter", trans. Geoffrey Winthrop-Young and Michael Wutz (Stanford: Stanford University Press, 1999; orig. pub. 1986), 2–3.
[37] Jerry Norman, "Chinese" (Cambridge: Cambridge University Press, 1988), 52–54.

ancient sounds, but a logical approximation and simulation of lost sounds using existing evidence.[38]

3.3 Visual Speech and Telephony

In the 19th century, the technological process of sound underwent a critical bifurcation and leap within the research of Alexander Graham Bell and his family. This family, deeply rooted in the historical background of deaf education, was impelled to explore two entirely new forms of sonic existence. First was *Visible Speech*, founded by his father Alexander Melville Bell. This was a phonetic notation system dedicated to translating the physiological movements of the vocal cords, tongue, and lips into visual symbols, representing a significant human attempt to image-encode the act of phonation itself rather than merely the acoustic result.[39] Subsequently, the younger Bell invented the telephone through the combination of acoustics and electricity. The Bell family inadvertently completed the technologies that transformed sound into images and electric currents; this not only changed the way sound is perceived but also completely detached it from the constraints of bodily presence and physical distance.[40]

3.4 The Phonograph

[38] William H. Baxter, "A Handbook of Old Chinese Phonology" (Berlin: Mouton de Gruyter, 1992), 27.
[39] Alexander Melville Bell, "Visible Speech: The Science of Universal Alphabetics" (London: Simpkin, Marshall & Co., 1867), 34.
[40] Jonathan Sterne, "The Audible Past: Cultural Origins of Sound Reproduction" (Durham: Duke University Press, 2003), 39–41.

Following the telecommunications era, the invention of the Phonograph by Thomas Edison and the earlier Phonautograph by Édouard-Léon Scott de Martinville marked a fundamental leap in the history of sound. For the first time in human history, instantaneous sound waves were physically captured and inscribed onto wax cylinders or discs. The phonograph dissolved the uncertainty and disposability of sound. Sound could be played back repeatedly and reproduced almost faithfully. This technology preserved not only linguistic meaning but also the speaker's timbre, breathing, and emotion.[41]

However, this also brought about the complete separation of sound from the vocal subject, namely the normalization of Acousmatic Sound. Since then, we have become accustomed to hearing voices without bodies.[42]

3.5 Medical Imaging of the Voice

According to all the subjects of my investigation mentioned above, we can see that traditional voice research is established on the foundation of Anatomy. However, anatomy presents a massive epistemological rupture: a cadaver is not a living body. Postmortem muscle tissue loses elasticity and begins to decay, unable to represent the state of life during phonation. To resolve this problem, starting from the invention of the laryngoscope by Manuel Garcia[43], to modern X-rays, MRI, and Stroboscopy, medical imaging technology has intervened in voice research.

---

[41] Friedrich A. Kittler, "Gramophone, Film, Typewriter", trans. Geoffrey Winthrop-Young and Michael Wutz (Stanford: Stanford University Press, 1999), 23–24.
[42] Michel Chion, "The Voice in Cinema", trans. Claudia Gorbman (New York: Columbia University Press, 1999), 18–19.
[43] Manuel Garcia, "Observations on the Human Voice," Proceedings of the Royal Society of London 7 (1855): 399–400.

The medical gaze has shifted; we have moved from observing static structures to observing Living Dynamics.[44] Medical imaging attempts to show us exactly how living bodily structures move. This technological intervention sparked my curiosity: Does our past understanding of sound based on cadaver dissection contain a massive cognitive deviation from the true state of life? Imaging technology allows us to "see through" the resonance and vibration inside the living body; sound is no longer mysterious but has become visible movement.

3.6 Synthetic Voice

In the related fields of sound and technology, currently, one cannot fail to mention Speech Synthesis. From the Voder machine of the 1930s[45] to modern AI based TTS (Text-to-Speech) technology, and on to virtual singers like *Hatsune Miku*[46]. We can see that through technological synthesis, humans can fabricate new human voices that have never existed before transcending the anthropocentric concept of the human body.

At this point, we can see the impact of technology on the human voice: writing stripped away the timbre of the voice, leaving only meaning; the phonograph preserved the timbre but stripped away the body; synthetic voice creates a sound with "no body, and not even an original speaker." In synthetic speech or the virtual personalities carrying speech, the voice is no longer the soul in the classical concept, but an algorithmic arrangement of

[44] See James Stark, "Bel Canto: A History of Vocal Pedagogy" (Toronto: University of Toronto Press, 1999), 6–8. For the foundational concept of the "medical gaze," see Michel Foucault, "The Birth of the Clinic: An Archaeology of Medical Perception", trans. A. M. Sheridan Smith (New York: Vintage Books, 1973), 89.
[45] Mara Mills, *Media and Prosthesis: The Vocoder, the Artificial Larynx, and the History of Signal Processing*, Quiet Dialogue 21 (2012): 107–149.
[46] Yuji Sone, "Japanese Robot Culture: Performance, Imagination, and Modernity" (New York: Palgrave Macmillan, 2017), 145–47.

data.[47] This poses ultimate ethical and ontological questions to us: When the voice can be perfectly calculated and generated, what is left of the "Humanity" that represents the "human" within the "Human Voice"?

[47] N. Katherine Hayles, "How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics" (Chicago: University of Chicago Press, 1999), 2–3.

Chapter 4 Thinking and the Intelligence from the Voice

Based on my personal artistic practice and research investigation, I believe that the human voice is the concrete embodiment of thought. As an output signal, the human voice embodies the fact that "thinking is taking place." It is precisely because of the human voice that language gains material form and is conveyed to the outside world through sound, realizing the exchange between the individual and the world. However, the role of the human voice is far more than just a carrier of information; it stands between invisible thought and tangible language, maintaining an intermediate state that is difficult to define simply.

This "intermediate state" reveals the profound connection between the voice and thought. Giorgio Agamben once said: "The search for the voice in language, this is what is called thought."[48] This statement suggests that the human voice is not merely an expressive tool, but a clue to the process of thought occurrence itself. However, does demonstrating the process of thought imply the establishment of intelligence? The process of thinking itself is not directly equivalent to the existence of intelligence. This leads to a questioning of my logical starting point: If the human voice is the connecting point between language and thought, then how exactly does it embody intelligence? What kind of symbiotic relationship is constituted between language and intelligence? Can we find the answer to the birth of intelligence within the mechanism of sound production?

To answer this question, we cannot rest on abstract definitions; we must examine the process of thinking from a physical level. After all, if thought cannot cross the boundary

[48] Giorgio Agamben, "Language and Death: The Place of Negativity", trans. Karen E. Pinkus and Michael Hardt (Minneapolis: University of Minnesota Press, 1991), 35.

from internal to external, intelligence is out of the question. This forces thought to seek a physical outlet. But this is not as simple as looking for accompanying phenomena. What we often call "internal thinking" is an internal activity lacking obvious external representation, essentially the active movement of neurons and signal transmission. To confirm and present this process, one must rely on the media. The most direct are speech, writing, and body movements, while more complex ones include public speaking, writing, and artistic creation. These media materialize thought itself, its generation process, and its methodology.

However, the relationship between different media and thought is not consistent. Writing or complex artistic creation are often products after thought is completed. They are relative static results of careful deliberation; before expression, the thought has largely taken shape. In contrast, spoken language and its accompanying real time body language are different. Speech possesses significant dynamism and immediacy.[49] When we speak, we are often displaying the process of thinking itself, rather than merely stating a presupposed result. The key to this difference lies in immediate feedback and correction. When you think while speaking, phonation and thinking occur almost synchronously. You may realize unclarity or logical deviation mid sentence, yet do not need to stop to delete and modify like in writing. Instead, you immediately correct the thought, and this correction is instantly reflected in the subsequent speech. When the brain organizes language and drives phonation, it is participating in the adjustment and correction of subsequent thought. This tight interactivity indicates that phonation is not just a unidirectional output of thought, but affects thought itself through a bidirectional action.

Considering that "Inner Speech" is often accompanied at the physiological level by suppressed "Subvocalization[50]," I assume that voice may point to the mechanism of

[49] Ong, "Orality and Literacy", 31–33.
[50] This perspective integrates Vygotsky's sociocultural theory with evidence from behavioral physiology. Vygotsky discussed the transformation of external speech into inner speech, L. S. Vygotsky, "Thinking and

intelligence generation. Specifically, "intelligence generation" here refers not to a static capacity or a preprogrammed set of rules, but to an emergent phenomenon arising from a continuous feedback loop[51]. In this context, the act of phonation functions as a cybernetic loop: the brain emits a signal, perceives the deviation through auditory or internal feedback, and instantaneously recalibrates the next cognitive move.[52] Therefore, intelligence is not a preexisting "source" of the voice, but rather the dynamic process of self correction triggered by the act of phonation itself. If we view "phonation" as this feedback mechanism for signal self correction, we approach the essence of intelligence. The production of thought stems from our ability to "hear" the signals we emit.

Then, does this feedback mechanism necessarily depend on the physical body we are familiar with?[53] Human trust in their own thinking is largely rooted in the carbon based flesh, because we can truly perceive the resonance of sound within the bodily structure. But if flesh is stripped away and replaced by silicon based chips and if the feedback loop is closed within a circuit rather than a larynx, does this trust still hold? This is no longer merely a discussion about whether artificial intelligence is "humanlike," but a fundamental inquiry into the attribution and the very birth of intelligence.

Speech", trans. Norris Minick (New York: Plenum Press, 1987; orig. pub. 1934); while subsequent research, particularly the electromyographic (EMG) experiments by A. N. Sokolov [A. N. Sokolov, "Inner Speech and Thought", trans. G. T. Onischenko (New York: Plenum Press, 1972)], conclusively demonstrated that inner speech is accompanied by faint motor impulses of the vocal organs, known as subvocalization.

[51] Karl Friston, *The Free-Energy Principle: A Rough Guide to the Brain?*, Nature Reviews Neuroscience 11, no. 2 (February 2010): 127–38.

[52] Norbert Wiener, "Cybernetics: Or Control and Communication in the Animal and the Machine" (Cambridge, MA: MIT Press, 1948).

[53] Andy Clark and David Chalmers, *The Extended Mind*, Analysis 58, no. 1 (January 1998): 7–19.

4.1 The Emergence of Intelligence and Ontological Commitment[54]

The logic underlying the linguistic expression of Artificial Narrow Intelligence (ANI)[55],[56] reveals certain limitations inherent in dialectics, prompting us to consider whether this stems from a lack of ontological commitment. This leads to a pivotal inquiry: does intelligence necessitate a steadfast adherence to ontology? Furthermore, does ontology necessarily presuppose the existence of a tangible entity, such as the human biological body, or the underpinning of a collective consciousness? How do humans formulate their ontology, and is it possible for them to forfeit this commitment entirely?

This divergence may well transcend the mere debate regarding the presence or absence of ontology, nor is it confined to a binary opposition concerning existence itself. My current assessment is that while intelligence may possess the potential to operate independently of ontology, for humans, ontological commitment is intrinsic to life itself; the physical body serves as the projection and symbol of this ontology within the material world.

Extrapolating further, even if we were to endow AI with a controllable physical form and sensory organs, assuming a simulation accuracy that infinitely approaches reality or physical capabilities, such as lens acuity, that surpass human limits, it would still fail to acquire truly human like embodied experience[57] so long as it lacks the organic integration and interplay between multiple senses. This fundamental deficit suggests that if AI were

---

[54] W. V. O. Quine, *On What There Is*, The Review of Metaphysics 2, no. 1 (September 1948): 32.
[55] Artificial Narrow Intelligence (ANI): Also referred to as "Weak AI," ANI designates AI systems programmed to perform a specific, singular task (e.g., facial recognition, language translation, or strategic games). Unlike Artificial General Intelligence (AGI), ANI operates under a limited set of constraints and lacks the cognitive ability to apply intelligence to tasks outside its predefined domain.
[56] Nick Bostrom, "Superintelligence: Paths, Dangers, Strategies" (Oxford: Oxford University Press, 2014), 22–26.
[57] Maurice Merleau-Ponty, "Phenomenology of Perception", trans. Donald A. Landes (London: Routledge, 2012; orig. pub. 1945), 84.

indeed to generate some form of "selfhood," it would necessarily represent a relationship between consciousness and the body grounded in a novel material basis,[58] radically distinct from that of carbon based life.

4.2 Blindsight and the Chinese Room

Following the preceding deduction, consciousness and intelligence can be conceptualized as two separable entities. This distinction transcends mere semantic categorization; rather, it is grounded in a fundamental misalignment between their physical actualization and philosophical essence. Within this framework, intelligence is reconstructed not as enigmatic intrinsic wisdom, but as a functional cybernetic loop—a dynamic error-correction mechanism predicated on signal emission and instantaneous calibration. Because this process is essentially a logic of deviation management, it possesses substrate independence; as long as the algorithm of phonation and correction is executed, intelligence is established, rendering it a transferable competence. In stark contrast, consciousness is anchored in an ontological commitment, emergent only from the unique resonance of organic embodiment. It is a state of being contingent upon the specific hardware of carbon based life, a non-transferable experience distinct from the software level logic of intelligence.

[58] Hayles, "How We Became Posthuman", 12.

The phenomenon of *blindsight* and the *Chinese Room* experiment[59],[60] serve as compelling evidence for this proposition. Blindsight refers to a condition where damage to the primary visual cortex results in the loss of subjective visual experience, yet the body remains capable of exhibiting unconscious responses to visual stimuli.[61] This demonstrates the existence of alternative processing pathways within the brain that do not generate subjective perception, thereby enabling the physiological structure to execute functions autonomously. Similarly, the Chinese Room thought experiment depicts an individual who, possessing no knowledge of the Chinese language, processes Chinese symbols solely by adhering to a rulebook to produce factually correct responses. From an external perspective, it appears as though understanding is present; however, the subject remains entirely ignorant of the semantic meaning of the symbols.[62]

Collectively, these examples underscore a critical insight: our physical constitution, particularly specific neural mechanisms, is capable of independently executing complex information processing and behavioral output. This process does not necessarily require the participation of consciousness.

This precipitates the central inquiry: Does understanding, as a cognitive state, inescapably demand the presence of subjective consciousness? Given that blindsight and the Chinese Room confirm the brain's capacity for complex, purposeful information processing

---

[59] John R. Searle, "Minds, Brains, and Programs", Behavioral and Brain Sciences 3, no. 3 (1980): 417.
[60] The Chinese Room Argument, which is a thought experiment proposed by John Searle to challenge the claims of "Strong AI" (the idea that a machine can truly understand language). Searle imagines a person who does not know Chinese locked in a room, following a set of English instructions (a computer program) to manipulate Chinese characters in response to input. To an outside observer, the person appears to understand Chinese; however, the person is merely performing syntactic symbol manipulation without any semantic understanding. In the context of Artificial Narrow Intelligence (ANI), this experiment is frequently cited to illustrate that AI's linguistic performance does not equate to genuine cognition or intentionality.
[61] Lawrence Weiskrantz, "Blindsight: A Case Study and Implications" (Oxford: Oxford University Press, 1986), 166.
[62] Searle, "Minds, Brains, and Programs," 417–18.

at an unconscious level, does this imply that consciousness is not the fundamental driving force behind the operation of intelligence?

Furthermore, if consciousness is not a diffuse function but rather a concrete physical entity, where exactly is it localized within the brain? Should science ultimately demonstrate the absence of an overarching consciousness center, would the unshakable ontological commitment constructed upon our sensory experiences be revealed as, in essence, a grandiose illusion resulting from the synchronous oscillation of countless neuronal firings?[63]

4.3 The Room without Echo

The logic of the Chinese Room experiment extends to the core of traditional computationalism, where the relationship between programmer, program, and computer is simplified into pure syntactic operations, essentially prioritizing linguistic logic. This experiment emphasizes that correct output is not equivalent to inner understanding, implying that information processing does not fully touch upon the essence of human intelligence. This directs our gaze toward the body, as physical structure actually undertakes many key functions we mistakenly attribute solely to the control of consciousness. This also validates Searle's critique of what he calls *Strong AI*,[64] rooted in his insistence on the biological basis of consciousness; he argues that perception is not an abstract algorithm, but an emergence of a biological entity's overall physical properties.[65]

---

[63] Francis Crick and Christof Koch, *Towards a Neurobiological Theory of Consciousness*, Seminars in the Neurosciences 2 (1990): 272.
[64] Searle, "Minds, Brains, and Programs," 420. Strong AI is a term coined by John Searle to describe the philosophical position that a computer running the right program would literally have a mind and consciousness, not merely simulate them. Proponents of Strong AI argue that mental states are purely computational processes (abstract algorithms) and can exist independently of the specific biochemical structure of the brain.
[65] Searle, "Minds, Brains, and Programs," 422.

Regarding this experiment, I attempt to propose a critique based on the perspective of voice. Searle deliberately chose unfamiliar Chinese symbols and a closed room to construct a pure symbol processing system ensuring the operator could not rely on a priori knowledge. Chinese symbolizes an absolute semantic rupture, while the room symbolizes sensory confinement. These two elements constitute an extreme metaphor for consciousness and the body: a closed entity cut off from external sensory interchange, operating solely on internal instructions. The narrowness of the room confirms that the person within can only perform internal computations, unable to perceive the greater external world.

However, I wish to press further: what exactly is the essence of that "rulebook"? Is it a directive, or a piece of sealed-off information? Even if the person in the room can consult and execute it, is this processing an active operation with intent, or an unconscious behavior akin to a physiological reaction? In my view, this handbook essentially interrupts the circuit of communication. The person in the room is like a perfect "aphasic" or "deaf" individual; he precisely casts out symbols yet never "reads" or "hears" the meaning of these symbols in the depths of his consciousness, implying that the dimension of voice is absent here.

Tracing back to the logic at the beginning of this chapter, if the emergence of intelligence stems from the synchronic cycle of "articulation → internal monitoring → correction → re-articulation,"[66] then the fundamental defect of the Chinese Room is that it is a cavity without resonance. Signals pass through the room without retention, failing to be transformed into internal sound. This is exactly like the predicament of blindsight: visual information passes through the brain triggering action, yet fails to stir the ripples of subjective awareness. This inspires me to re-examine the so-called ontological commitment. Perhaps it refers to the necessity for a system—whether a human brain or artificial intelligence—to possess a "cavity" that allows information to generate an echo. Whether

[66] Willem J. M. Levelt, "Speaking: From Intention to Articulation" (Cambridge, MA: MIT Press, 1989), 13–14.

this cavity is woven from neurons or constructed by code is perhaps not the most critical factor; the true key lies in whether it can "hear" itself.

Here, I do not view voice as some form of abstract rhetoric; on the contrary, it should be established as an indispensable substantial intermediary in the operation of intelligence. Voice is essentially a physical action that forcefully bridges inner hidden thought with the outer objective world. It is not merely a unidirectional output of information but constructs an immediate mechanism of Information Reflux: thought is transformed into sound waves through physical vibration, projected externally, and immediately captured by auditory or internal senses, feeding back to the center of consciousness. This self-referential loop of output as input enables the thinker to "hear" and correct their own conscious activity in real-time at the very moment of expression.

It is precisely this mechanism that profoundly reveals the essence of ontological commitment: the generation of intelligence must depend on a physical anchor capable of sustaining this cycle of internal and external oscillation. Whether this anchor is the flesh and blood of a carbon-based organism or some future precision silicon carrier, the key lies in whether it possesses the structural capability to generate a flow of information between "inner—outer—inner." Only by possessing this perceptible feedback loop can intelligence transcend cold calculation and truly acquire a sense of presence in the world.

Conclusion

Guided by the "vocal organ" as a core thread, this study constructs a unique "Apparatus for Generating Voice and Images" and, through three stages of artistic practice and interdisciplinary theoretical examination, deeply explores the complex entanglement between voice, body, and intelligence from physical, anatomical, technical, and philosophical perspectives. Initially stemming from a microscopic observation of my mother reconstructing her perceptual modes amidst hearing loss, this research eventually expands into a macroscopic inquiry into the ontology of art and "embodiment" in a post-human context.

In terms of practice, this study verifies through three projects that the body is not a passive vessel for voice but an active constructor of cognition. Project I, through researching the isomorphism between "body and instrument," reveals how humans utilize external tools (such as the pipe organ and Peruvian whistling vessels) to externalize and simulate their own vocal mechanisms. Project II, through an anatomical examination of the "hyoid bone and articulatory organs," demonstrates that vocalization is not a localized mechanical vibration but a systemic balance of forces throughout the body; it also critiques the discipline and alienation of the body within the history of Audism. Project III, by constructing a "vocal tract model" and discovering the logic of "spatial inversion" in casting, transforms the invisible negative space of the oral cavity into a visible, tangible entity, thereby confirming the field of voice's existence on a material level.

In terms of theory, this paper first examines the rupture between the "signifier" and the "signified" from a linguistic perspective. The study points out that in traditional linguistics, voice is often viewed as a transparent conduit leading to the "signified (meaning)," while its

own materiality is frequently overlooked. However, with the development of media technology, voice has undergone an evolution from the silent archiving of text, to the "separation of voice and flesh" realized by the phonograph, and finally to the complete stripping of the subject by contemporary AI synthesized speech. In this process, the materiality of voice has continuously turned from solid to void, ultimately reduced to dimensionless pure information within algorithms.

Philosophically, this technological evolution implies a loss of "Presence"—voice is no longer the confirmation of a subject's existence but has degenerated into a data stream that can be infinitely replicated by algorithms, posing a fundamental challenge to the philosophical status of the "human voice." Through a re-examination of the "Chinese Room" thought experiment and the phenomenon of "Blindsight," this paper points out that the reason AI cannot generate true intelligence is precisely because it is trapped in a game of pure symbolic computation (signifiers). It stops at the syntactic processing of linguistic symbols, lacking a sensory entity that anchors these symbols to the real world, thus failing to achieve an embodied understanding of semantics. Within this horizon, vocalization is not merely a carrier of language but a vibration originating from within the biological being—it is the key to breaking the "prison of symbols." It forces the subject to touch the world through the vibration of the flesh, establishing a physical loop of "sending signal—receiving feedback" (hearing oneself), thereby endowing language with the weight of the body. Therefore, vocalization constitutes an "Ontological Commitment": there must exist an entity capable of producing an "echo" to complete the confirmation of its own existence in the qualitative leap from "symbol" to "meaning." This is precisely the boundary that distinguishes true consciousness from simulated intelligence.

In summary, although modern technology can accurately simulate the acoustic characteristics of voice and even achieve the separation of voice from flesh at the algorithmic level, this has not touched the essence of the vocalizing act. As a physical action that confirms the existence of thought, the core value of vocalization lies in constructing a feedback loop that connects the internal and the external. My artistic practice, those clay and mechanical devices simulating the vocal tract, does not aim to replicate the perfect human voice, but attempts to reenact the process of my mother learning to speak, seeking new logic of connection amidst absence and limitation. This seemingly primitive and heavy simulation process is, in essence, a repair of perception alienated by technology. It responds to the philosophical proposition regarding machine consciousness: true intelligence cannot stop at symbolic computation; only when a machine possesses a physical space that allows it to perceive its own vibrations, thereby establishing an ontological commitment to its own existence, can it truly walk out of that room devoid of echoes. Here, art is no longer merely an aesthetic practice but evolves into a methodology that changes in real-time with reality. As artificial intelligence increasingly approaches human cognition, returning to the body (returning to that structural entity capable of resonance) may be the key to understanding the irreplaceable nature of humanity, consciousness, and artistic creativity.

Bibliography

Agamben, Giorgio. "Language and Death: The Place of Negativity." Translated by Karen E. Pinkus and Michael Hardt. Minneapolis: University of Minnesota Press, 1991.

Arensburg, Baruch, Anne-Marie Tillier, Bernard Vandermeersch, Henri Duday, Lynne A. Schepartz, and Yoel Rak. *A Middle Palaeolithic Human Hyoid Bone*. Nature 338, no. 6218 (April 1989): 758–60. https://doi.org/10.1038/338758a0.

Baxter, William H. "A Handbook of Old Chinese Phonology." Berlin: Mouton de Gruyter, 1992.

Baynton, Douglas C. "Forbidden Signs: American Culture and the Campaign against Sign Language." Chicago: University of Chicago Press, 1996.

Bell, Alexander Melville. "Visible Speech: The Science of Universal Alphabetics." London: Simpkin, Marshall & Co., 1867.

Bostrom, Nick. "Superintelligence: Paths, Dangers, Strategies." Oxford: Oxford University Press, 2014.

Bradbury, Jack W., and Sandra L. Vehrencamp. "Principles of Animal Communication." 2nd ed. Sunderland, MA: Sinauer Associates, 2011.

Chion, Michel. "The Voice in Cinema." Translated by Claudia Gorbman. New York: Columbia University Press, 1999.

Clark, Andy, and David Chalmers. *The Extended Mind*. Analysis 58, no. 1 (January 1998): 7–19.

Crick, Francis, and Christof Koch. *Towards a Neurobiological Theory of Consciousness*. Seminars in the Neurosciences 2 (1990): 263–75.

Derrida, Jacques. "Of Grammatology." Translated by Gayatri Chakravorty Spivak. Baltimore: Johns Hopkins University Press, 1976. Originally published 1967.

Dolar, Mladen. "A Voice and Nothing More." Cambridge, MA: The MIT Press, 2006.

Double-Whistle Vessel with Sea Birds. Sicán (Lambayeque) culture, northern Peru. 900–1100. Earthenware. The Walters Art Museum, Baltimore. Anonymous gift, 2009. Accession no. 48.2826.

Evans, Dayna. *Pharmakon, Sound Art, and Expressing Bodily Disturbance Through Noise.* Pitchfork, October 23, 2014. https://pitchfork.com/thepitch/556-pharmakon-sound-art-and-expressing-bodily-disturbance-through-noise/.

Fay, Richard R., and Arthur N. Popper. *Evolution of Hearing in Vertebrates: The Inner Ears and Processing.* Hearing Research 149, no. 1–2 (2000): 1–10.

Fisher, Elizabeth, Daniel Austin, Heather M. Werner, and Elizabeth Gebhard. *Hyoid Bone Fusion and Bone Density Across the Lifespan: Prediction of Age and Sex.* Forensic Science International 262 (2016): 282.e1–282.e6.

Foucault, Michel. "The Birth of the Clinic: An Archaeology of Medical Perception." Translated by A. M. Sheridan Smith. New York: Vintage Books, 1973.

Friston, Karl. *The Free-Energy Principle: A Rough Guide to the Brain?* Nature Reviews Neuroscience 11, no. 2 (February 2010): 127–38. https://doi.org/10.1038/nrn2787.

Garcia, Manuel. *Observations on the Human Voice.* Proceedings of the Royal Society of London 7 (1855): 399–410.

Gray, Henry. "Anatomy of the Human Body." 20th ed. Edited by Warren H. Lewis. Philadelphia: Lea & Febiger, 1918.

Halliday, David, Robert Resnick, and Jearl Walker. "Fundamentals of Physics." 10th ed. Hoboken, NJ: John Wiley & Sons, 2013.

Hayles, N. Katherine. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago: University of Chicago Press, 1999.

Holland, John. "The Mark and the Voice." In The Clinic of Transference, edited by Colette Soler, Paris: Éditions du Champ lacanien, 2004, 147–155.

Hoobler, Ellen, and Sonia Matheus. *The Art of Knowing When Not to Wet Your Whistle: Multidisciplinary Research into the Use of a Sicán Whistling Vessel in the Museum Context.* The Walters Journal 75 (2017): 1–20.

Hornbostel, Erich Moritz von, and Curt Sachs. *Systematik der Musikinstrumente: Ein Versuch*. Zeitschrift für Ethnologie 46, no. 4–5 (1914): 553–590.

Jin, Jie. "Chinese Music." Cambridge: Cambridge University Press, 2011.

Kempelen, Wolfgang von. "Mechanismus der menschlichen Sprache / The Mechanism of Human Speech: Kommentierte Transliteration & Übertragung ins Englische / Commented Transliteration & Translation into English." Edited by Fabian Brackhane, Richard Sproat, and Jürgen Trouvain. Dresden: TUDpress, 2017.

Kittler, Friedrich A. "Gramophone, Film, Typewriter." Translated by Geoffrey Winthrop-Young and Michael Wutz. Stanford: Stanford University Press, 1999. Originally published 1986.

Kooijman, Piet G., Felix I. de Jong, Geert Thomas, and Wim Huinck. C*omparative Study of the Relation Between Laryngeal and Body Posture in Patients with Muscle Tension Dysphonia and Controls.* Journal of Voice 19, no. 3 (2005): 366–374.

Krause, Bernie. "The Great Animal Orchestra: Finding the Origins of Music in the World's Wild Places." New York: Little, Brown and Company, 2012.

Lacan, Jacques. "Écrits: The First Complete Edition in English." Translated by Bruce Fink. New York: W. W. Norton & Company, 2006.

Lane, Harlan. "When the Mind Hears: A History of the Deaf." New York: Random House, 1984.

Levelt, Willem J. M. "Speaking: From Intention to Articulation." Cambridge, MA: MIT Press, 1989.

Lieberman, Philip. *The Evolution of Human Speech: Its Anatomical and Neural Bases.* Current Anthropology 48, no. 1 (2007): 39–66.

Merleau-Ponty, Maurice. "Phenomenology of Perception." Translated by Donald A. Landes. London: Routledge, 2012. Originally published 1945.

Mills, Mara. *Media and Prosthesis: The Vocoder, the Artificial Larynx, and the History of Signal Processing.* Quiet Dialogue 21 (2011): 24–49.

Myers, Thomas W. "Anatomy Trains: Myofascial Meridians for Manual and Movement Therapists." 4th ed. London: Elsevier, 2020.

Namiki, Chizuru, Koji Hara, Haruka Tohara, Kohei Kobayashi, Ariya Chantaramanee, Kazuharu Nakagawa, Takuma Saitou, et al. *Tongue-Pressure Resistance Training Improves Tongue and Suprahyoid Muscle Functions Simultaneously.* Clinical Interventions in Aging 14 (2019): 601–8. https://doi.org/10.2147/CIA.S194808.

Norman, Jerry. "Chinese." Cambridge: Cambridge University Press, 1988.

Ong, Walter J. "Orality and Literacy: The Technologizing of the Word." 2nd ed. London: Routledge, 2002.

Pettit, Nathan J., and Ronald C. Auvenshine. *Change of Hyoid Bone Position in Patients Treated for and Resolved of Myofascial Pain.* CRANIO 38, no. 5 (2020): 283–90.

Pijanowski, Bryan C., Luis J. Villanueva-Rivera, Sarah L. Dumyahn, Almo Farina, Bernie L. Krause, Brian M. Napoletano, Stuart H. Gage, and Nadia Pieretti. *Soundscape Ecology: The Science of Sound in the Landscape.* BioScience 61, no. 3 (March 2011): 203–216. https://doi.org/10.1525/bio.2011.61.3.6.

Plato. "Phaedrus." Translated by Alexander Nehamas and Paul Woodruff. Indianapolis: Hackett Publishing, 1995.

Quine, Willard Van Orman. *On What There Is.* The Review of Metaphysics 2, no. 1 (September 1948): 21–38.

Riches, Martin. "The Talking Machine." Martin Riches. Accessed December 5, 2025. https://martinriches.de/talk.html.

Sachs, Curt. "The History of Musical Instruments." New York: W. W. Norton & Company, 1940.

Saussure, Ferdinand de. "Course in General Linguistics." Edited by Charles Bally and Albert Sechehaye. Translated by Wade Baskin. New York: The Philosophical Library, 1959.

Searle, John R. *Minds, Brains, and Programs.* Behavioral and Brain Sciences 3, no. 3 (1980): 417–57.

Sokolov, Aleksandr Nikolaevich. "Inner Speech and Thought." Translated by G. T. Onischenko. New York: Plenum Press, 1972.

Sone, Yuji. "Japanese Robot Culture: Performance, Imagination, and Modernity." New York: Palgrave Macmillan, 2017.

Standring, Susan, ed. "Gray's Anatomy: The Anatomical Basis of Clinical Practice." 42nd ed. London: Elsevier, 2020.

Stark, James. "Bel Canto: A History of Vocal Pedagogy." Toronto: University of Toronto Press, 1999.

Statnekov, Daniel K. "Animated Earth: A Story of Peruvian Whistles and Transformation." 2nd ed. Berkeley, CA: North Atlantic Books, 2003.

Sterne, Jonathan. "The Audible Past: Cultural Origins of Sound Reproduction." Durham: Duke University Press, 2003.

Titze, Ingo R. "Principles of Voice Production." Iowa City: National Center for Voice and Speech, 2000.

Weiskrantz, Lawrence. "Blindsight: A Case Study and Implications." Oxford: Oxford University Press, 1986.

Wiener, Norbert. "Cybernetics: Or Control and Communication in the Animal and the Machine." Cambridge, MA: MIT Press, 1948.

Winzer, Margret A. "The History of Special Education: From Isolation to Integration." Washington, DC: Gallaudet University Press, 1993.

Wolfe, Joe. "How do woodwind instruments work?" UNSW Music Acoustics. Accessed January 14, 2026. https://www.phys.unsw.edu.au/jw/woodwind.html.

Wu, Katherine J. *This Tiny Fossil Shows How Our Ears Evolved From Reptilian Jaws.* NOVA, PBS. June 6, 2019.

Yale, Caroline "A. Formation and Development of Elementary English Sounds." Northampton, MA: Gazette Printing Co., 1892.