SONIC HYPERMIRROR: ATTUNING TO HYPEROBJECTS

by

Wonki Kang

Bachelor of Architecture
Seoul National University, 2016

Submitted to the Department of Architecture and
The Department of Electrical Engineering
in Partial Fulfillment of the Requirements for the Degrees of

Master of Science in Architecture Studies and

Master of Science in Electrical Engineering and Computer Science

at the

Massachusetts Institute of Technology

February 2022

Signature of Author:____
Department of Architecture
Department of Electrical Engineering
January 14, 2022

Certified by:____
Axel Kilian
Visiting Professor of Architecture
Thesis Supervisor

Accepted by:____
Leslie K. Norford
Professor of Building Technology
Chair, Department Committee on Graduate Students

Accepted by:____
Leslie A. Kolodziejski
Professor of Electrical Engineering
Chair, Department Committee on Graduate Students

Thesis Supervisor:
Axel Kilian, PhD
Visiting Professor of Architecture

/

Thesis Reader:
Arvind Satyanarayan, PhD
Assistant Professor of Computer Science

SONIC HYPERMIRROR: ATTUNING TO HYPEROBJECTS

by

Wonki Kang

Submitted to the Department of Architecture and
the Department of Electrical Engineering and Computer Science on

January 14, 2022

in Partial Fulfillment of the Requirements for the Degrees of

Master of Science in Architecture Studies and
Master of Science in Electrical Engineering and Computer Science

ABSTRACT

Long-running pandemic without an end in sight, climate crisis encroaching on our everyday lives—global crises are collective events, but they take on multiple forms and scales, leading to radically different experiences for people. The inter-scalar, inter-temporal representations gained dire urgency due to the crises surfacing simultaneously at a global scale. *Hyperobject*, as defined by ecological philosopher Timothy Morton, is the in-experienceable object that is vastly distributed in time and space that easily exceeds human's perceptive capability. I start with a hypothesis: *hyperobjects are better heard than seen*. This thesis is focused on the critical approach to data representation, by bringing forward listening as a primary modality of interaction. I present Sonic Hypermirror, a custom tool that allows data probing of large-scale audio data based on vocal interaction, accompanied by a visual interface that utilizes computational tools to assemble a soft, continuous semantic space of multiple audio streams. It is an experiment to build a data sensorium where the listeners enter into, inhabit, and learn from. Through the thesis, I propose the system of data representation that is continuous, non-referential, and exploratory; and revisit the affordances of architectural space as a data storage and an interactive datascape.

Thesis Supervisor:
Axel Kilian, PhD
Visiting Professor of Architecture

Acknowledgments

I was incredibly privileged to be at MIT where I was welcomed to learn from, inspired by, and collaborated with so many amazing minds. This thesis would have not been possible without them. I express my gratitude to:

Arvind Satyanarayan for teaching me that any form of representations involves politics behind it.

Landon Morrison for helping me navigate the field of sound studies and teaching me how to listen to the world differently.

Axel Kilian, for your persistent caring and support, intellectually and emotionally. Thank you for giving me the courage to march forward to the murky path of interdisciplinary studies.

+

My pandemic friends *Wu, Elina, Catherine, Semine,* and *Mar* for reminding me that I am not alone living through this unprecedented, challenging time.

My fellow ex-pats *Jinu* and *Howon,* who were of incredible help coping with cultural transition, and to Kiwon and Miru for making sure I eat healthy foods during the difficult time.

Jonah, for showing me that there is still so much to explore in this world, and for inspiring and supporting me to finish this work.

Last but not least, to mom and dad, and my sister *Suyeon*.

Prologue

Before I moved to the United States for my graduate study in 2019, I lived in apartments commonly found in the city of Seoul, rows of concrete boxes built during the breakneck-paced modernization in the post-war period. The apartment complex was usually completely segregated from the life happening outside of the small, isolated universe, and the concrete enclosures were thick enough to muffle the noises throbbing from busy streets.

When I first moved to a century-old wood frame house in Cambridge, it took some time for me to get used to being able to feel the vibration borne through the building structure—squeaking of the hardwood floor from the neighbors climbing up and down the stairs, repetitive trembling from the washing machine, and the nerve-racking quivering from the windstorms. The influx of information traveling through the architectural structure was felt through and within my body, and architecture, became an extension of my bodily senses.

When the pandemic struck in the spring of 2020, there was *silence*. The air in my room was still, but it was saturated with radiowaves and electric signals, which connected me to the outside world. I started to think, how could I regain physical contact to the outside world, and other bodies—what are the ways to distance myself from others physically but not socially?

Table of Contents

Sonic Hypermirror: Attuning to Hyperobjects

by

Wonki Kang

Words, in so many ways, falter and fail. One can only talk about fear, sorrow, and exhaustion in a language reduced to its ossified minimal components, a language that has reached the edge of what it can express, approximating silence but not falling into it just yet.[1]

— Rosi Braidotti

[1] "'We' Are In This Together, But We Are Not One and the Same," *Journal of Bioethical Inquiry* 17, no. 4 (December 2020): 465–69.

Figure 1. The San Francisco Bay area obscured in orange smoke and haze due to a massive wildfire in 2020. (Photograph by Christopher Michel)

01 Introduction

1.1. Attuning to Hyperobjects

The year 2020 and 2021 have brought a vast range of long-running global crises to light. Climate change is more felt than ever, and the pressures of the global pandemic further escalated racial injustice. Trying to make sense of such phenomena is overwhelming, and the efforts of journalists to explain them in words or visual representations have repeatedly failed to evoke empathy from people. Global crises are collective events, but they take on multiple forms and scales, leading to radically different experiences for people.

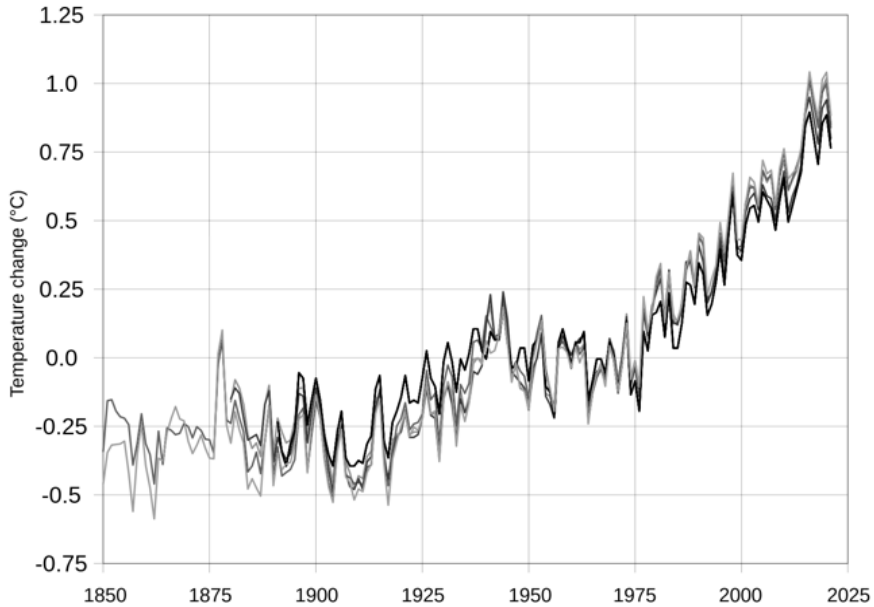


Figure 2. Annual global surface temperature, 1850-2020.
Data from NASA Goddard Institute for Space Studies, Berkeley Earth,
and National Oceanic and Atmospheric Administration.

Experiences of the crises are highly subjective and are dependent on the person's social, economic, and even physiological conditions. While this divergence of experiences is unavoidable, an inter-scalar, inter-temporal representation to provide a common language of communication gained dire urgency due to the crises surfacing simultaneously at a global scale.

The challenge is that events arising at a global scale are hyperobjects, as defined by ecological theorist Timothy Morton, objects that are in-experienceable, for they are vastly distributed in time and space beyond human's perceptive capability.[2] Hyperobjects inevitably involve radically different temporality and scale than the ones relatable to humans.

---

[2] Timothy Morton, Hyperobjects: Philosophy and Ecology after the End of the World (University of Minnesota Press, 2013).

In 2020, the count of wildfires in California set a record of close to ten thousand, burning more than four million acres of land. There have been debates over the cause of the increasing cases of wildfires especially in California, but climate change is not-so-arguably one of the most significant causes of the anomaly. The orange sky caused by the fires created an otherworldly yet formidable spectacle throughout the major cities in California, resembling that of the apocalyptic depiction of the future in the science fiction films like *Blade Runner* (Figure 1). While this atmospheric phenomenon widely raised people's awareness of climate change, the weather is a "local representation of a much larger phenomenon that is strictly imperceptible."[3]

Abstract representations are seemingly better at giving a complete view of a large-scale phenomenon. (Figure 2) However, the rational description of complex phenomena often falls short of emotional engagements to the object. Furthermore, the supposed objectivity of the data visualization has been debunked by scholars arguing that data visualization is biased, thus showing a narrow perspective but is presented as "truth." How can we close the gap between the psychological distance of abstraction and the visceral engagement to things?

Central to the thesis is the questioning of how hyperobjects can be made experienceable. For hyperobject cannot be known directly or holistically, Morton suggests *attuning* to it.[4] Attunement describes an act to "tune or bring into harmony," as musical instruments, or to tune in to a certain frequency of signals as a radio, and even expands to describe building a deep connection with someone.[5] How can we go about *attuning* to the things that we are not even sure that they exist? How do we relate to the ghostly monster—one feels that it is there but not quite sure if it is real?

---

[3] Ibid, 87.
[4] Timothy Morton, "Attunement," *Ecology Without Nature* (blog), November 2, 2014, http://ecologywithoutnature.blogspot.com/2014/11/ attunement.html.
[5] *Merriam-Webster Dictionary*, s.v. "attune," accessed November 4, 2021, https://www.merriam-webster.com/dictionary/attune.

The concept of attunement directly relates to composer Pauline Oliveros's idea of *deep listening*. As one of the central figures in the development of post-war experimental music, she defines deep listening as a listening practice where the listeners are invited to challenge their habitual understandings and constantly break down the boundaries of the box of preconceptions. In one of her talks little more than a year before her passing, she beautifully describes the concept of deep listening as:

> learning to expand perception of sounds to include the whole space-time continuum of sound, encountering vastness and complexities as much as possible … one is connected to the whole of the environment and beyond."[6]

Attuning represents a "heightened state of awareness and connects to all that there is," allowing you to open yourself up to transform and be transformed through listening.[7] To attune to others denotes listening to their minds, which is by Hogan's definition, *empathy*.[8]

1.2. Listening to Others

Launched in 2020, an up-and-coming social media application *Clubhouse* exploded in its number of users throughout 2020 to 2021. Conversations in the platform are in real-time and thus ephemeral, and the community guideline states that they strictly regulate any form of redistribution. Here is a short excerpt of the business statement from the founders of the platform:

---

[6] TEDx Talks, *The Difference between Hearing and Listening | Pauline Oliveros | TEDxIndianapolis*, 2015, https://www.youtube.com/watch?v=_QHfOuRrJB8.
[7] Pauline Oliveros, "Quantum Listening: From Practice to Theory (To Practice Practice)," 2000.
[8] Robert Hogan, "Development of an Empathy Scale," *Journal of Consulting and Clinical Psychology* 33, no. 3 (1969): 307–16.

> "Clubhouse is voice-only, and we think voice is a very special medium ... The intonation, inflection and emotion conveyed through voice allow you to pick up on nuance and form uniquely human connections with others. You can still challenge each other and have tough conversations—but with voice there is often an ability to build more empathy. This is what drew us to the medium."[9]

With the digital platform synchronizing the users onto the same time continuum, I was able to fully tune in to the social interaction that is *taking place* in the digital space, which allowed me to build a deeper connection than any other social media platforms. Furthermore, vocal interactions freed the virtual identities of the users from the bias that would have existed on other forms of visual-based social media.

Roland Barthes distinguishes between hearing and listening—hearing is a physiological phenomenon, whereas listening is a psychological act.[10] Actively listening, or *attuning* to others, not to judge or to analyze, but to embrace and to connect, is the key to empathy.

While the listening experience in Clubhouse is mostly on trying to understand the meanings of human speech, we can expand this observation beyond human voice. Michel Chion defines three modes of listening: *causal, codal*, and *reduced listening*. Causal listening, the most common, is to discern the source of the sound; codal listening is to extract meanings of sound, primarily practiced in listening to human speech; and finally, reduced listening is to take sound as an independent entity free from its symbolic property.[11] These distinctions in listening offer opportunities to speculate on what opportunities lie in between the three modes. In the following paragraphs, I briefly map out the relationships between the three modes and try

---

[9] Paul Davison and Rohan Seth, "Clubhouse—Check 1, 2, 3... Is This Thing On?," Clubhouse, accessed April 23, 2021, https://www.joinclubhouse.com/check-1-2-3.
[10] Roland Barthes, "Listening," in *The Responsibility of Forms: Critical Essays on Music, Art, and Representation*, 1st ed. (New York: Hill and Wang, 1985), 245–60.
[11] Michel Chion, "The Three Listening Modes," in *Audio-Vision : Sound on Screen* (New York: Columbia University Press, 1994), 22–34.

to answer what are the challenges and opportunities that arise from the comparisons.

Does reduced listening separate the listeners from the context? Acousmatic situations make the listeners to actively search for the ground of the sound: they can exacerbate causal listening because it deprives us of the aid of other sensory modalities.[12] In other words, the acousmatic evoke *imagination*—it urges the listeners to construct an idea of materiality of the objects that might have caused the sound, the speaker's virtual identity imagined from their voice, or even an image of a place that the sound might have been originated from.

Is codal listening limited strictly to linguistic sounds? The distinction between language and non-language based on their "meaningfulness" undermines human capability of learning. Mark Johnson argues that "meaning is more than words and deeper than concepts," reinforcing the significance of the *body* in meaning-making, and expanding the notion of language beyond semantics.[13] Vocal bursts that seemingly have no meaning sometimes deliver even more meanings than speech.

Going back to the question of hyperobjects, they approach you in myriad ways — from the material to immaterial, artificial to natural, temporal to timeless, liquid to solid, and so on. Jumping forward to concluding to a singular interpretation of them can be dangerous, preventing us from actively expanding our notion of the ever-changing nature of hyperobject. Pivoting between the three modes of listening, we can start to attune to hyperobjects rather than reducing them into a box of symbolic languages.

[12] Ibid, 28.
[13] Mark Johnson, *The Meaning of the Body: Aesthetics of Human Understanding* (Chicago, IL: University of Chicago Press, 2008).

1.3. Music as Environment, Environment as Music

One might argue that musical listening is qualitatively different from everyday listening. Everyday listening, listening to events that lack any structure that a composer would have prescribed for musical pieces, involves discerning the sound events that are spectrally messy. On the other hand, musical listening involves understanding the harmonicity and motifs. However, the conceptual distinction between musical listening and everyday listening has been radically obscured by experiments carried out by musicians and artists since the mid-20th century.

Composer Halim El-Dabh (1921 - 2017) in his early 20s, sneaks into the Zār exorcism ceremony, cunningly dressed as a woman, with a magnetic tape recorder in his hand, recorded the ambient sound of the strictly female exclusive ritual.[14] The piece *The Expression of Zār*, is known as the very first piece of *musique concrète*, predating the most commonly known, Pierre Schaffer's.[15] El-Dabh's fascination with the possibilities of manipulating tape-recorded sounds to create music continued throughout his career. El-dabh moves to New York City in the 1950s. Surrounded by the noises of the metropolitan city, he would take screeching sounds from the subway trains and experiment by physically chopping the magnetic tapes and splicing the pieces.[16]

Four years later after the *Expression of Zār,* Pierre Schaeffer (1910 - 1995) in Paris wrote his first piece of musique concrete *Étude Aux Chemins de Fer (Study of the Railways),* an acoustic montage created from a palette sampled from sounds pro-

[14] Tommy McCutchon, "Unlimited Americana: A Conversation with Halim El-Dabh," Music & Literature, June 2017, https://www.musicandliterature.org/features/2017/6/1/unlimited-americana-a-conversation-with-halim-el-dabh.
[15] Sam Shalabi and Halim el-Dabh, "Step into the Electric Magnet: Music Pioneer Halim El-Dabh," *Bidoun: Arts and Culture from the Middle East* 1 / 9 (January 1, 2007): 78–80.
[16] Denise A. Seachrist, Halim El-Dabh, and George Faddoul, *The Musical World of Halim El-Dabh.* (Kent, Ohio : Kent State University Press, c2003., 2003), 61.

duced from trains like metal grinding from their wheels, train whistles, and air-horns. Seemingly banal or even off-putting sounds took on unique characters as they were combined, processed, and filtered in creative ways. Schaeffer describes his personal experience working with musique concrète: "I was looking for the opposite of music but it came back on me violently."[17]

The endeavors to rediscover the musicality of sound and to reunify music, noise, and language, gave rise to hyper-musicality—music as environment, the environment becoming music in its entirety, in which Maryanne Amacher's words, "the split which now exists between these two worlds—that of musical language and of environmental sound" is being closed.[18] Then, we now can use sound as a building block of constructing aural architecture; or the other way around, "space [being] the grain of sound."[19]

---

[17] Jean de Reydellet, "Pierre Schaeffer, 1910-1995: The Founder of 'Musique Concrète,'" *Computer Music Journal* 20, no. 2 (1996): 10–11.

[18] Maryanne Amacher, "Proposal for Anywhere city to National Endowment for the Arts (NEA)'s City Options Program" in *Maryanne Amacher: Selected Writings and Interviews* (originally 1974, Brooklyn, NY: Blank Forms Editions, 2021).

[19] Jonathan Sterne, "The Stereophonic Spaces of Soundscape," in *Living Stereo: Histories and Cultures of Multichannel Sound* (New York: Bloomsbury Academic, 2015), 65–83.

02 Background

2.1. Aural Architecture

In 1967, a group of architects, painters, musicians, and sculptors at Yale School of Art and Architecture founded Pulsa. Their statement *The City as an Artwork* was published in György Kepes's book *Arts of the Environment* begins with their diagnosis of how the dichotomy between architecture and information system are becoming more and more blurry, saying that "the information systems are becoming more architectural, while the architecture is becoming less object-like and more systemic."[20] They were deeply invested in sound as it enables "softening" of the hard surfaces and its ability to transform the architectural and urban spaces.

[20] The Pulsa Group, "The City as an Artwork," in *Arts of the Environment*, 1972, http://archive.org/details/TheCityAsAnArtwork, 208.

Aural architecture is architecture made of sound, and architecture of sound. Aural architecture involves the relationships of the space beyond the physical properties and acoustics of the space, encapsulating both emotional and behavioral dimensions of architecture.

During the 1950s, Halim El-Dabh composed pieces with hours of tape recordings, processing and rearranging them using various techniques including re-recording in a room with operable walls to create different acoustic qualities.[21]

Alvin Lucier's paradigmatic work *I am Sitting in a Room*, first recorded in 1969, was an experiment that questioned the relationship between the physical space and the sound that occupies it. Throughout the repeated steps of playing a tape recording of himself narrating a text in the room and re-recording the voice, frequencies that do not resonate with the frequencies in the room get attenuated, creating a gradual merge between the acoustic quality of the room and the voice[22]—a loop that oscillates between sound occupying the space, and the space being made in sound.

William Gaver explains sound in relation to our physical environment in the context of the theory of interaction: "sound provides information about an interaction of materials at a location in an environment."[23] Spatial awareness experienced through sound is more than merely "informative", but able to affect our emotional and behaviors.

Going further, sound is not only "informative" of the environment, but it creates "place." Cultural historian Josh Kun calls this place in between real and virtual, *audiotopia*, where he explains the auditory experience not only as our reception of stimulus through the vibration of our eardrums but as to experience "a space we

---

[21] Thom Holmes, *Electronic and Experimental Music: Technology, Music, and Culture*, 3rd ed. (New York: Routledge, 2008).
[22] "DRAM: Notes for 'Alvin Lucier: I Am Sitting in a Room,'" accessed January 20, 2022, https://www.dramonline.org/albums/alvin-lucier-i-am-sitting-in-a-room/notes.
[23] William W Gaver, "What in the World Do We Hear? An Ecological Approach to Auditory Event Perception," *Ecological Psychology* 5, no. 1 (1993): 19.

enter into, encounter, move around in, inhabit, be safe in, [and] learn from."[24] Sound transports us to a different world through our inner experience.

2.2. Evocative Objects

The unruly relationship of the sign and the signified is one of the most notable characteristics of audition, which poses both challenge and opportunity simultaneously. Does this ambiguity of sound conjure evocation?

The term "evocative object" was popularized by technology theorist Sherry Turkle, along with her book *Evocative Objects: Things We Think With* published in 2007. She describes evocative objects in the context of the radical proliferation of telecommunication technology, as "objects [that are] companions to our emotional lives or ... provocations to thought," objects that mediate between individuals and the bigger world.[25]

2.2.1. Thematic Apperception Test

In the 1930s, artist Christiana Morgan and clinical psychologist Henry Murray at Harvard developed Thematic Apperception Test (TAT), a psychological test to gauge subjects' personalities, needs, fantasies, emotions, and their disorders. The subjects were asked to imagine stories from a set of pictures—what is the present situation, the characters in the picture thinking and feeling, and what story would have led up to the story. For example, TAT picture 3BM (Figure 3), which is considered one of the most important pictures in the set, illustrates a boy crouching with an ambiguous object next to him. This particular picture was known to

---

[24] Josh Kun, "Strangers among Sounds," in *Audiotopia: Music, Race, and America*, American Crossroads 18 (Berkeley: University of California Press, 2005), 2.
[25] Sherry Turkle, *Evocative Objects: Things We Think With* (Cambridge, Mass: MIT Press, 2007).

be used to diagnose personalities regarding the subject's relationship with their family, and their behavioral disorders; for instance, one of the criteria is whether the patient sees themselves as a parent in the figure; also whether they describe the ambiguous object as a weapon, etc.[26] Morgan and Murray, when they were designing the images, they wanted them to be "evocative objects," which they would describe as objects into which a subject "project his circumstances, experiences or preoccupations."[27]

Figure 3. Thematic Apperception Test picture, code 3BM.

[26] Silvans S. Tomkins, *The Thematic Apperception Test: The Theory And Technique Of Interpretation*, 1947, 109.
[27] Mara Mills, "Evocative Object: Auditory Inkblot," *Continent* 5, no. 1 (2016): 15–23.

What followed visual TAT was an aural version of the projective test: auditory apperception test. In 1953, clinical psychologists Thomas Ball and Louis Bernardoni designed a test consisting of thirteen sound effects: mechanical devices, musical instruments, crowds, non-verbal human sounds, animals, and sounds from nature. Subjects were then asked to "put the sounds together to make up a dramatic story" upon listening to the audio.[28]

2.2.2. Ambiguity and Evocation

The ambiguity of the TAT pictures allowed them to be filtered in specific and telling ways to which the subjects were made to fill in the "missing parts" with their own interpretation.[29] Human-computer-interaction theorist William Gaver advocates ambiguity as a "resource for design" that can make interaction evocative rather than didactic. For Gaver, ambiguity is a property of the interpretative relationship between people and artifacts. Compared to the related concepts like fuzziness or inconsistency that are attributes of things, ambiguity is an attribute of our interpretation of them.[30] It is worth mentioning that Braverman and Chevigny noted that "vision tends to provide more continuing cues, or reality checks, than does audition"[31]—in other words, audition tends to put the subject in a position where they could be more tolerable being delivered fictional events without putting them into the mode of skepticism. Ambiguity alone is not enough to resonate with the human mind. We need to be able to trust what we are sensing to allow ourselves with the pluralism of our perception. Again, the practice of listening

---

[28] Thomas S. Ball and L. C. Bernardoni, "Application of an Auditory Apperception Test to Clinical Diagnosis.," *Journal of Clinical Psychology* 9 (January 1953): 54–58.
[29] Mills, "Evocative Object: Auditory Inkblot."
[30] William W. Gaver, Jacob Beaver, and Steve Benford, "Ambiguity as a Resource for Design," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03 (New York, NY, USA: Association for Computing Machinery, 2003), 233–40.
[31] Sydell Braverman and Hector Chevigny, *The Braverman-Chevigny Auditory Projective Test* (American Foundation for the Blind, Inc., 1951).

without the source of the sound being seen is called acousmatic or reduced listening—to mute the non-aural modalities free oneself from the expectations and schema that potentially interfere with the listening.[32]

2.2.3. The Myth of the Empathy Machine

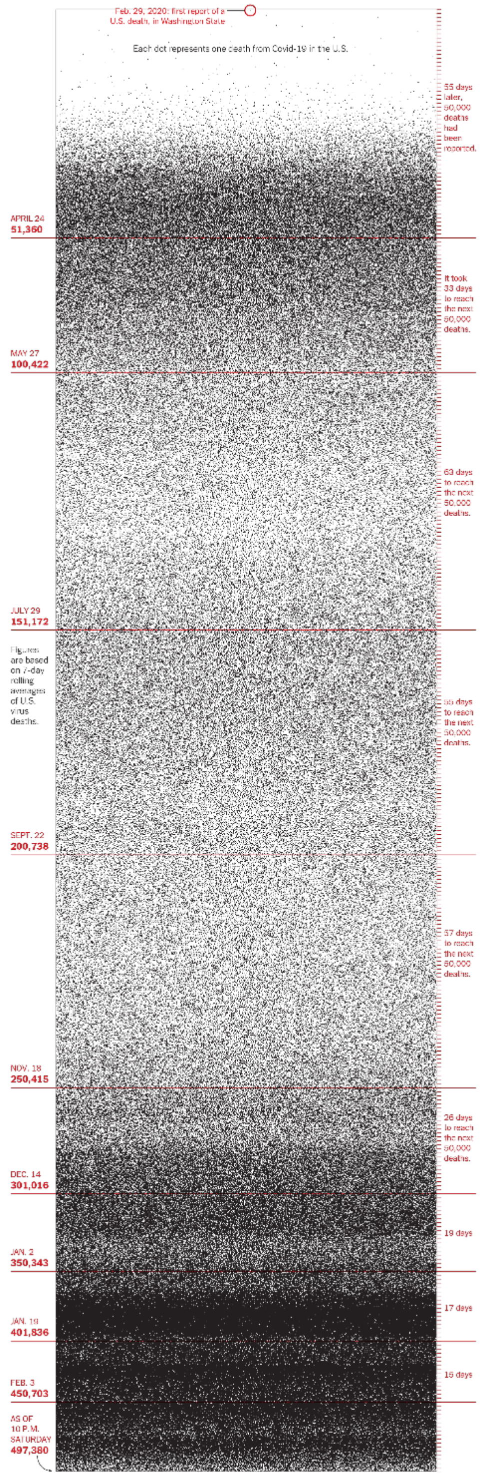
An American film critic Roger Ebert in one of his interviews, says that "the movies are like a machine that generates empathy."[33] The term "empathy machine" was echoed by an artist and filmmaker Chris Milk, arguing that Virtual Reality (VR) is an "ultimate empathy machine."[34]

However, the empathy machine argument is a double-sided sword. Blind pursuit of empathy, an attempt to clone and distribute the experiences in their rawest form possible, is alarming in that it neglects the interpretative relationship between the subject and the object, which is in fact one of the most crucial elements of empathy. To empathize is to engage with others emotionally, and it is achieved by blurring the boundary between the subject and the object. While the media technology can help bring the object to the subject as close and vivid as possible, the question that has yet to be answered is to do with whether the subjects are willing to transform themselves to feel the others. Empathy cannot be "injected" through the VR goggles, and the experiences cannot be transmitted from one person to another through a conduit. Immersiveness does guarantee empathy; it is the "god trick of seeing everywhere from nowhere."[35] In the flood of images, we are rather becoming more and more blind.

---

[32] David Beard, "Acousmatic Listening and a Critical Awareness of Place," *International Journal of Listening* 33, no. 3 (September 2019): 129–32.
[33] *Roger Ebert on Empathy*, accessed January 12, 2022, https://www.rogerebert.com/empathy/video-roger-ebert-on-empathy.
[34] Chris Milk, *How Virtual Reality Can Create the Ultimate Empathy Machine*, 2015, https://www.ted.com/talks/chris_milk_how_virtual_reality_can_create_the_ultimate_empathy_machine.
[35] Donna Haraway, "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective," *Feminist Studies* 14, no. 3 (1988): 575.

Figure 4. "The Toll: America Approaches Half a Million Covid Deaths" from The New York Times issue of Februrary 21st 2021



2.3. Seeing Beyond Vision

2.3.1. The Limitations of Data Visualization

"The Toll: America Approaches Half a Million Covid Deaths" from *The New York Times* illustrates the number of deaths from Covid-19 in the United States as it was approaching half a million.[36] (Figure 4) The deaths are represented as dots, where denser dots shade the canvas darker, creating an abstract yet evocative image that expresses a "tide of deaths." However, it is not free from the unease that derives from the fact that the death of the individuals is reduced into a small bl ack dot. The nature of the visual representation is reductive—the more one tries to compress more information to the canvas, the more abstract the visualization becomes.

[36] Nancy Coleman, "On the Front Page, a Wall of Grief," *The New York Times*, February 21, 2021, sec. Times Insider, https://www.nytimes.com/2021/02/21/insider/covid-500k-front-page.html.

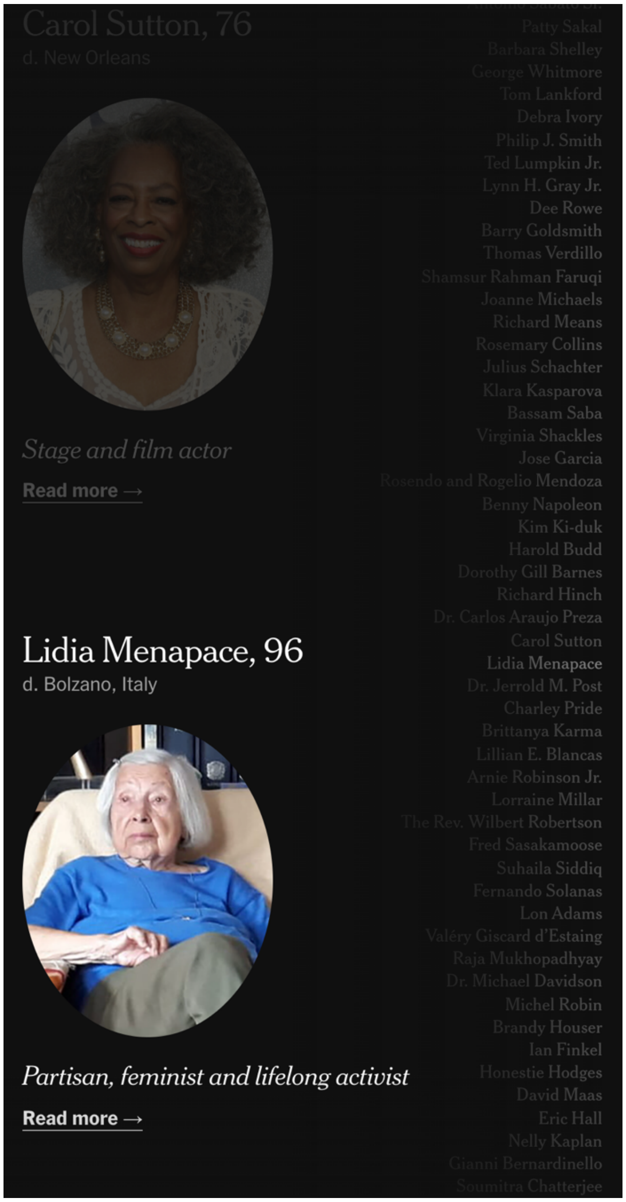


Figure 5. "Those We've Lost", an Interactive Obituaries of Covid-19 Deaths from the New York Times, April 16, 2020 - ongoing

Figure 5 is another example that portrays Covid-19 related deaths, an ongoing online obituary of people who died from Covid-19 with their names and portraits.[37] The editor of the project Daniel J. Wakin notes that the goal of the project "was never to give a comprehensive accounting of Covid-19's death toll, but to put at least some faces on the quickly multiplying numbers."[38] Compared to Figure 4, while not giving a "comprehensive" view of the Covid-19 related death, discovering the name and face of a neighbor, a friend, or someone you admired does induce emotional and visceral reaction to the viewers.

It is worth noting that data visualization has long been seen as neutral and objective, as designer Mushon Zer-Aviv would call "the unempathetic art."[39] Attempts for emotional engagements in data visualization have long been seen as embellishments, labeled "unscientific." However, the debate over the objectivity of representation is at stake.

---

[37] Daniel J. Walkin, "Those We've Lost," *The New York Times*, April 16, 2020, sec. Obituaries, https://www.nytimes.com/interactive/2020/obituaries/people-died-coronavirus-obituaries.html.t
[38] Daniel J. Wakin, "'Those We've Lost,' a Chronicle of Covid Death, Comes to a Halt," *The New York Times*, June 4, 2021, sec. Times Insider, https://www.nytimes.com/2021/06/04/insider/covid-obituaries-those-lost.html.
[39] Mushon Zer-Aviv, "DataViz—The UnEmpathetic Art," *Responsible Data* (blog), October 19, 2015, https://responsibledata.io/2015/10/19/dataviz-the-unempathetic-art/.

Donna Haraway was among the first who discussed the underlying partiality of data visualization, stating that one can only have a partial perspective of the data through their (mostly visual) representations. In the article *Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective*, she terms the effect of visualizations to be presented as an exhaustive description of data as the god trick—they trick the viewers to believe that they are looking from above, seeing everything.[40] She also argued against the predominance of vision in the practice of representing data by saying that the visual "technologies have bolstered claims to scientific objectivity and fortified uneven distributions of power."[41]

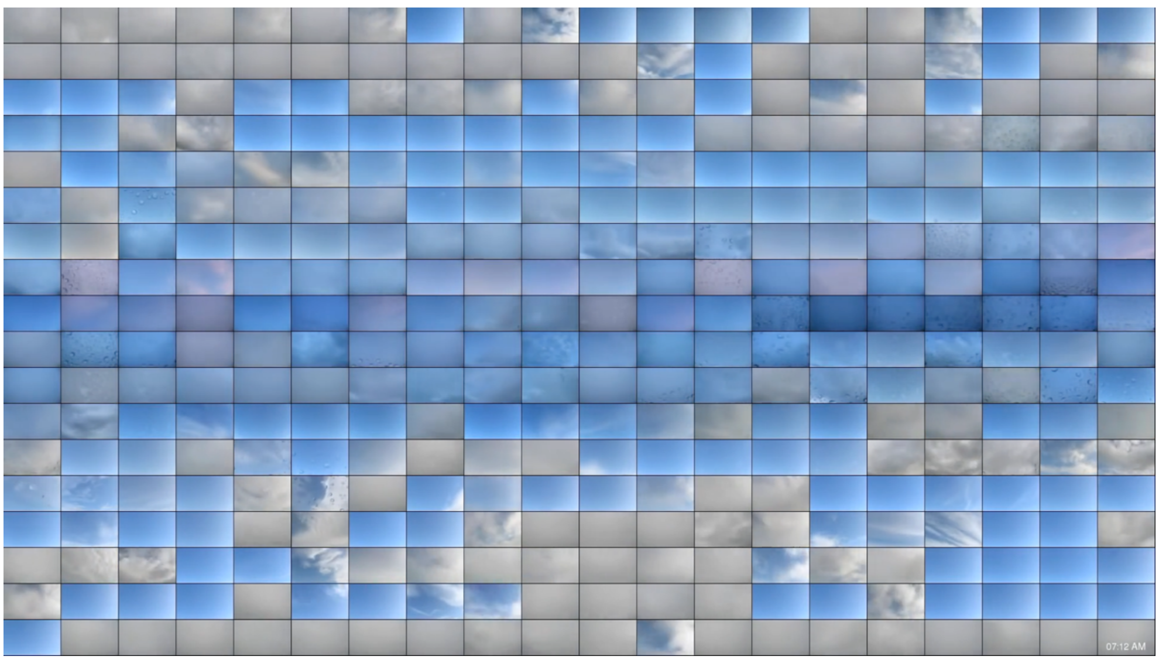


Figure 6. A History of the Sky by Ken Murphy

[40] Haraway, "Situated Knowledges."
[41] Tara Rodgers, "Approaching Sound," in *The Routledge Companion to Media Studies and Digital Humanities*, ed. Jentery Sayers, 1st ed. (New York : Routledge, Taylor & Francis Group, 2018.: Routledge, 2018), 233–42.

A History of the Sky by artist Ken Murphy is an example using a visual medium to represent parallels of time. A year-long video recording taken from a camera in a fixed position is compressed into a four-minute video fast-forwarded and subdivided into multiple points in time.[42] The limitation of the work lies in the fact that it is limited to a single geographical location, offering a single perspective rather than inviting the viewers into a sphere of experience. Jonathan Sterne writes, "hearing immerses its subject, vision offers a perspective."[43] Vision always has its strict boundaries exemplified by the grid layout in this example, and the question is how we can incorporate not just inter-scalar or inter-temporal, but spatio-temporal quality in its entirety.

Can sound overcome the "partial perspective" and potentially offer the "close-to-god" experience?

2.3.2. Data Sonification

Voegelin explains this nature of sound and listening as a framework from which to "interrogate the surface of a visual world."

> ... the world of sound is an event world, in contrast to that of vision which is an object world (Ong, 1971): it is a world of activities rather than artefacts, sensations rather than reflections (Schäfer, 1985). It is dynamic: something is happening for sound to exist. (Pocock, 1989: 193)[44]
>
> — Salomé Voegelin

In the data science community, there has been a growing interest in data sonification, a sonic representation of data, hoping the inference of the data is better preserved, meaning that the perceptual distance between the data in their raw form

[42] Ken Murphy, *A History of the Sky*, 2011, https://www.youtube.com/watch?v=PNln_me-XjI.
[43] Jonathan Sterne, "Sonic Imaginations," in *The Sound Studies Reader* (New York: Routledge, 2012), 9.
[44] Salomé Voegelin, *Sonic Possible Worlds: Hearing the Continuum of Sound* (New York London New Delhi Sydney: Bloomsbury, 2014).

and sonified is closer than it is in vision. This cross-domain translation is mainly achieved by translating the parameters from the data directly into the parameters of soundwaves. In the human-computer interaction community, the sonic medium has been chiefly studied pragmatically, thus strictly symbolic, such as the design of auditory icons or alarms.[45]

Figure 7. Numbers Station by Mendi and Keith Obadike (2015 – present)

Sound art duo Mendi and Keith Obadike's work shows an alternative of symbolic use of sound and the tradeoff intrinsic in visual representation. In their piece Numbers Station (2015-present), they "sonify" the data of the stop-question-and-frisk program in New York City by reading the numbers from NYPD reports with their voices.[46] (Figure 7) This meta-representation of the data, a re-embodiment of numbers assembled through the human voice, an aggregate index of racialized violence, manifests second lives.

[45] See William Gaver, "Auditory Icons: Using Sound in Computer Interfaces," *Human-Computer Interaction* 2, no. 2 (June 1, 1986): 167–77.
[46] Mendi and Keith Obadike, *Numbers Station 1 [Furtive Movements]*, 2016, https://youtu.be/PuLzv53gM_o.

2.4. Big Data, Big Spaces, and Imagination

Nonlocality is one of the major characteristics of hyperobject. Hyperobjects are manifested through distributed events more substantially than the local ones. Telecommunication technology enabled transmitting sensory experience from one place to another giving rise to the possibility of overcoming geographical constraints in human experience.

2.4.1. City

At 7 p.m. on March 27, 2020, hundreds of New Yorkers leaned out from their apartment windows and balconies and came in the unity of cheers. Hashtag *#ClapBecauseWeCare* circulated social media creating a movement to recognize and show gratitude toward essential workers on the frontlines of the fight against the COVID-19 crisis (Figure 8). This collective symphony of cheering, clapping, banging pots and pans brought the people together within a grand sphere of time-space in the scale of the metropolitan city. This noisy practice as a means of activism dates to at least the 19th century when French women banged pots and pans to protest economic conditions and food shortages. *Cacerolazo* refers to the "sonic protests" that gained prominence in the 1970s in Chile and rippled throughout Latin America.[47]

[47] Ruby Mellen, "How Pots and Pans Became Tools of Protests, from Chile to Myanmar," *Washington Post*, February 4, 2021, https://www.washingtonpost.com/world/2021/02/04/pots-pans-protests-myanmar-coup/.

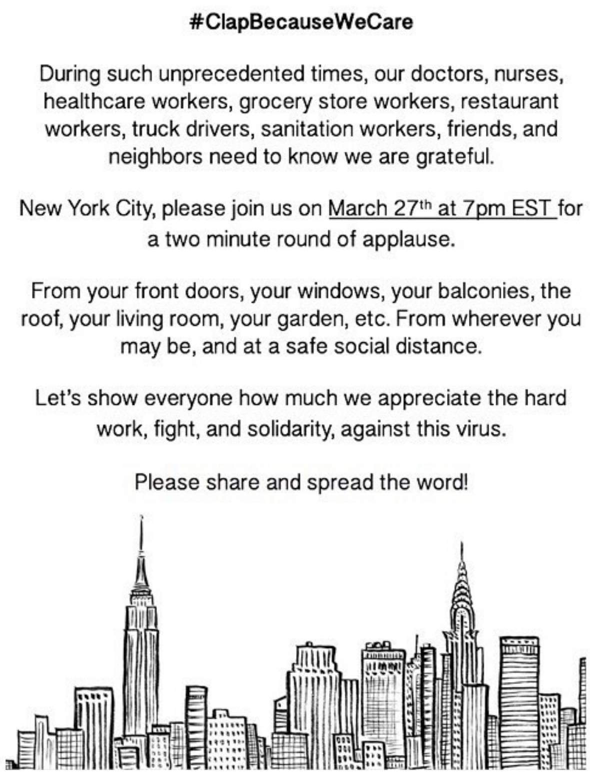


Figure 8. Hashtag #ClapBecauseWeCare circulated in New York City.
source: Instagram @gillllly

This simple but powerful anecdote shows the potential of sound creating a shared space that offers the sense of "being there together at the same time," an aural architecture beyond architectural scale.

Composer Maryanne Amacher (1938 - 2009) aspired to conjoin geographically distant places through sound. Her long-term project "Long Distance Music" was a series of text prompts to overcome the physical distance and bring different locations under the same sphere of music.

2.3.2. Earth

> [I'm interested in sounds] in which exaggerates naturally occurring, all organic sounds and phenomena, materials that exist in the moment […] like a piano that's mountain-size high and imagining what it would sound like…[48]
>
> — SOPHIE (Sophie Xeon)

On the morning of the 27th of August 1883, an earth-shattering sound blasted from a volcanic eruption in the island of Krakatoa, Indonesia. This otherworldly sound was loud enough to injure eardrums from 100 miles and could be fully heard 2,000 miles from Krakatoa. A report states that people were hearing "explosions resembling blasting of a rock" from regions in South Australia.[49] This unprecedented natural disaster brought almost a thirteenth of the globe into a single sphere of perception. Pauline Oliveros defines the concept of *sonosphere*: "sonosphere is the sonorous or sonic envelope of the earth."

For listeners with enough listening experiences, their sensory perception is sophisticated enough to discern different sounds in extremely high granularity. For example, sonic experience of water—a single waterdrop, a water stream, and then the ocean—listeners can immediately discern the scale of certain sonic events. Can we draw an analogy of the collective clapping as the ocean of claps?

---

[48] Arte TRACKS, *SOPHIE: The Producer Taking Pop to the Future (English Version / Interview) | Arte TRACKS*, 2018, https://www.youtube.com/watch?v=2ifh0tDrwBA.
[49] Sir Richard Strachey, "The Air Waves and Sounds Caused by the Eruption of Krakatoa in August, 1883," in *The Eruption of Krakatoa: And Subsequent Phenomena* (Trübner & Company, 1888), 57–88.

2.3.3. Room

Now we are back in our room again, with a pair of speakers wired to our computers. While the air is still, it is, in fact, saturated with radiowaves and electric signals. We enter the stereophonic space through which the speaker drums vibrate from electric signals construct, coordinated by the machines. The surfaces of the building enclosure are blurred and transformed, as we enter the space of an unknown scale through our imagination mediated by the digital signals.

The mirror is, after all, a utopia, since it is a placeless place. In the mirror, I see myself there where I am not, in an unreal, virtual space that opens up behind the surface; I am over there, there where I am not … From the standpoint of the mirror I discover my absence from the place where I am since I see myself over there.[50]

— Michel Foucault

[50] "Of Other Spaces," trans. Jay Miskowiec, *Diacritics* 16, no. 1 (1986): 22–27.

03 Building Digital Aural Architecture

3.1. Building Blocks: Units of Sound

3.1.1. Overview: Representing Sound

It is crucial to prescribe ways of representing sound carefully and purposefully along with defining the unit of sound as it sets the very foundation for further steps of search and synthesis. There lies a long history of symbolic representation, exemplified by the musical scores designed for the purpose of communicating composers' intention to the performers. The piece could be "realized" only during it is performed.

The rift between the symbolic and the "real" started to develop from the point where sounds could be captured to the fidelity that resembles the reality brought about by the invention of the gramophone. Representations of sound before gramophone were symbolic and instructive for the yet-to-be-realized. The invention of the gramophone in the 19th century marked the epistemological shift of sound from symbolic representation to the ossified reality; as Friedrich Kittler states, “the real takes the place of the symbolic.”[51] Gramophone enabled what musicologist Mark Katz would call *performative quotation* in sampling, “quotation that recreates all the details of timbre and timing that evoke and identify a unique sound event.”[52] Now with all the reality-capturing apparatus readily available to the public, defining the unit of sound is analogous to defining the unit of sonic reality, to be more precise, defining the gestalt of our auditory perception.

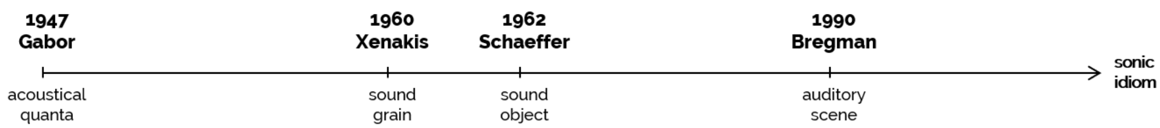


Figure 9. A brief history of the units of sound

Mathematician Dennis Gabor was the first to define the unit of sound in terms of human perception. In his paper *Acoustical Quanta and the Theory of Hearing* published in *Nature* back in 1947, he defines the lower bound in defining the size of the unit of sound, a “quantum” of hearing, which takes into consideration both time and frequency pattern in our auditory sensation.[53] From the post-war era, scientific discoveries and theorizations in signal processing paved way for innovations in the musical composition community.

While composer Iannis Xenakis borrows a similar concept as Gabor, he was more focused on sampling natural sounds and chopping them into *grains*, rearranging them to compose music, which gave rise to the concept of granular synthesis. Pierre Schaeffer was more interested in sounds from the perspective of listeners

[51] Friedrich A. Kittler, *Gramophone, Film, Typewriter*, trans. Michael Wutz and Geoffrey Winthrop-Young (Stanford: Stanford University Press, 1999).
[52] Mark Katz, “Music in 0s and 1s: The Art and Politics of Digital Sampling,” in *Capturing Sound: How Technology Has Changed Music* (University of California Press, 2004)., 149.
[53] D. Gabor, “Acoustical Quanta and the Theory of Hearing,” *Nature* 159, no. 4044 (May 1947): 591–94.

as opposed to the perspective of musical composers. Centered around the Groupe de Recherche Musicale (GRM), he conducted experiments exploring the possibilities of recorded sound materials as a resource for musical composition.[54] In his seminal piece *Traité des Objets Musicaux* (*Treatise on Musical Objects*), Schaeffer defines the notion of *sound object*, a segment in a source recording that sets the basic unit of composition.[55]

Cognitive scientist Bregman defines the term "auditory scene" that focuses more on the perception of a listener, how the human auditory system organizes sound into perceptually meaningful elements.[56] In contrast to vision, sonic perception tends to cluster certain qualities from a continuous stream of information. Similar definitions include Temporal Semiotic Unit (TSU).[57]

Going back to the three modes of listening, pivoting between the three modes encourages crossing disciplines among musicians, cognitive scientists, linguists, and so on. For example, Cowen, Elfenbein, Laukka, and Keltner created a 2-dimensional map of vocal bursts that seemingly do not have linguistic property but expresses a variety of emotions. (Figure 10) Computational tools allow the complex sounds to be projected onto a continuous semantic space to classify them based on the emotions expressed through human vocalizations.[58] In their work, there is no distinguishment between a musical note, language, or a component in a soundscape. The non-speech vocalizations represent non-referential meanings on their own.

---

[54] Reydellet, Jean de. "Pierre Schaeffer, 1910-1995: The Founder of 'Musique Concrète.'" *Computer Music Journal* 20, no. 2 (1996): 10–11.
[55] Diemo Schwarz, "Concatenative Sound Synthesis: The Early Years," *Journal of New Music Research* 35 (March 1, 2006): 3–22.
[56] Albert S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound* (Cambridge: The MIT Press, c1990).
[57] François Delalande, "Les Unités Sémiotiques Temporelles: Problématique et Essai de Definition," in *Les Unités Sémiotiques Temporelles* (Marseille: MIM, 1996).
[58] Alan Cowen et al., "Mapping 24 Emotions Conveyed by Brief Human Vocalization," *American Psychologist*, December 20, 2018.

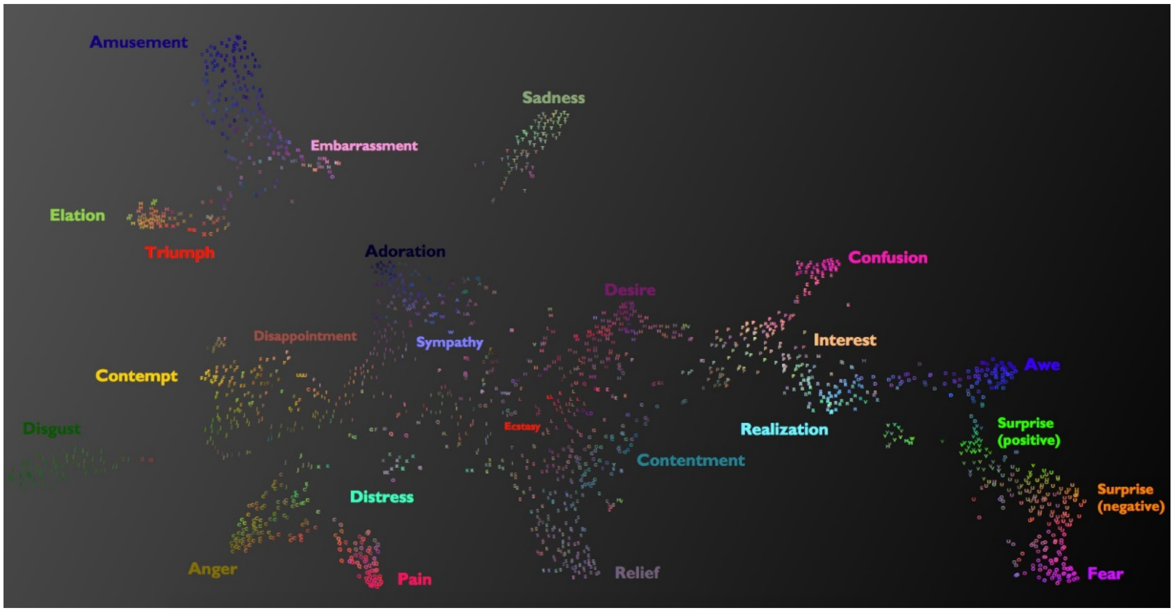


Figure 10. 24 Emotions Conveyed by Brief Human Vocalization. credit: Cowen et al.

3.1.2. Towards Sonic Idioms

3.1.2.2. Defining Sonic Idiom

In the thesis, "sonic idiom" is proposed as a unit that sets the basis for my sound composition. According to *Cambridge Dictionary*, idiom is defined as "a group of words whose meaning considered as a unit is different from the meanings of each word considered separately." Sonic idiom is a high-level, highly flexible sonic unit, a chunk of audio that contains the meaning that binds to it. Sonic idiom is a material to write a sonic poem. The idiom can be as short as a less-than-a-second metallic clink or as lengthy as a whole music track.

In the seminal book *Metaphors We Live By*, cognitive linguists George Lakoff and Mark Johnson articulated their concept of structural metaphors. Also known as conceptual metaphor, structural metaphor is a metaphor in which one concept is understood and expressed in terms of other concepts.[59] They argue that metaphors are deeply ingrained in human learning. Meanings stick to other concepts, creating a nexus of knowledge that is expanded, pruned, and blended. Fauconnier and Turner further extend the concept of metaphor, focusing more on the "blending" of conceptual spaces, or semiotic spaces.[60] Goguen also articulated his concept of algebraic semiotics, an approach to semiotics that can allow us to describe the structure of signs or symbols in a mathematical manner.[61]

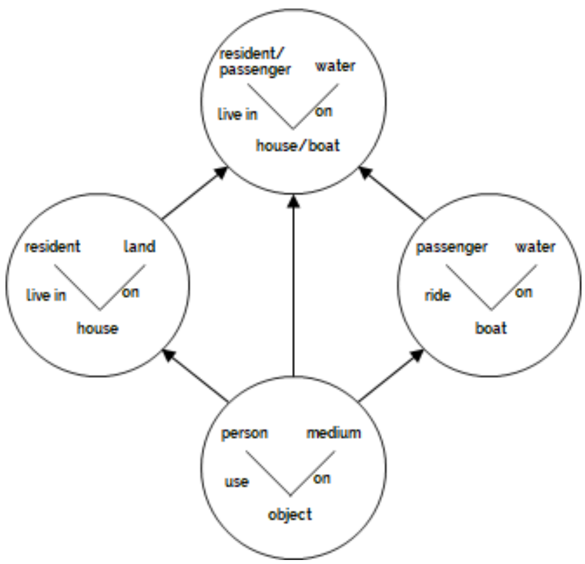


Figure 11. Diagram explaining how the new concept of "houseboat" emerges through the conceptual blending of "house" and "boat"

[59] George Lakoff and Mark Johnson, *Metaphors We Live By* (Chicago: University of Chicago Press, 1980).
[60] Gilles Fauconnier and Mark Turner, *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities* (New York, NY: Basic Books, 2002).
[61] Joseph Goguen, "An Introduction to Algebraic Semiotics, with Application to User Interface Design," in *Computation for Metaphors, Analogy, and Agents*, ed. Chrystopher L. Nehaniv, vol. 1562, Lecture Notes in Computer Science (Berlin, Heidelberg: Springer Berlin Heidelberg, 1999), 242–91.

We can apply the idea of conceptual metaphor to sound composition. "Sonic metaphor" is constructed through the conceptual blending of the sounds we are already familiar with. Listening is highly contextual. Analogous sounds can be interpreted in an entirely different way depending on the context or the listeners' cultural background. Also, as the audio stream starts to create a rhythm or be combined with different sound objects, it starts to construct more complex and affective meanings.

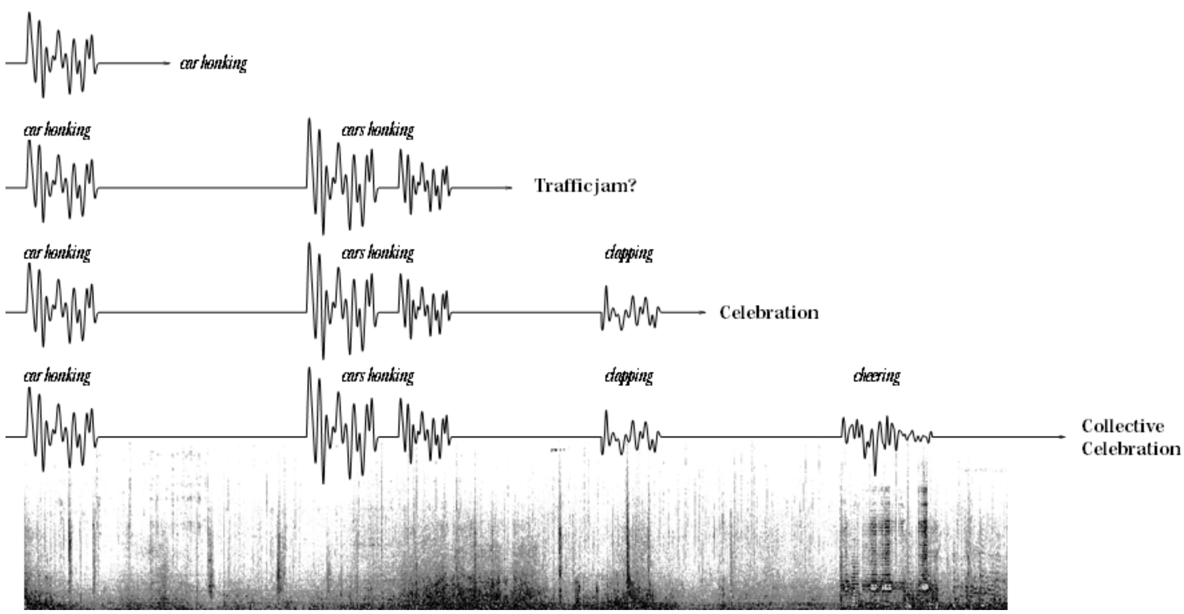


Figure 12. Unfolding of the temporal meaning in listening

Let's listen to an example of a short audio footage recorded on November 7th, 2020, at Harvard square area to make sense of how we make meaning as the audio unfolds through time. It was the day when the 2020 presidential election result was announced, and the street was full of people celebrating the result: cheering, cars honking, and singing. The sound of a car honking alone does not invoke any complex affective meaning, but multiple cars honking can be interpreted as a traffic jam or a sort of celebratory event. Then as you start to hear the clapping of people on the street in the background, the idea of the celebratory event is reinforced, making it clearer that it is a celebration that takes place in the series of sonic events. (Figure 12)

3.1.2.2. Concatenating Idioms

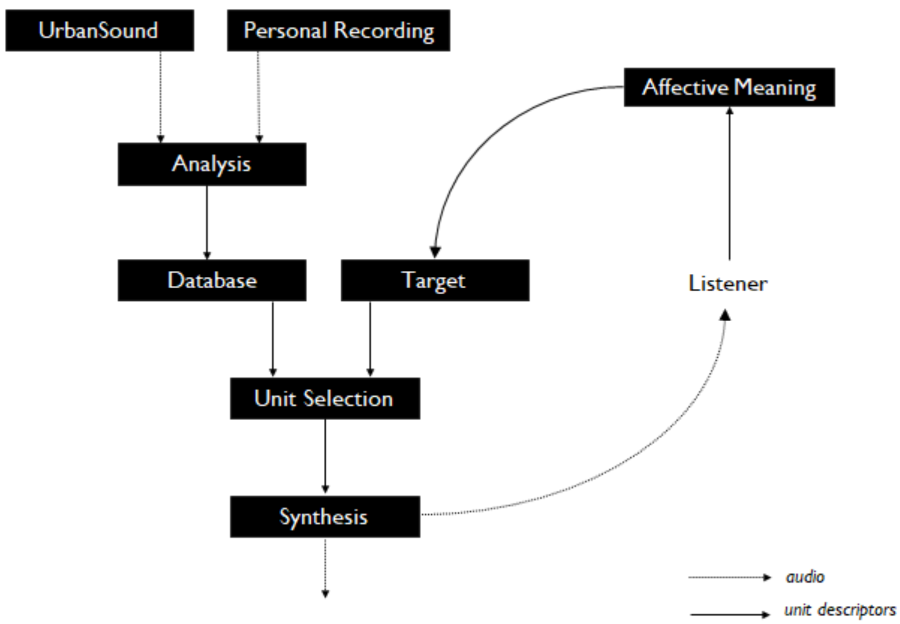


Figure 13. Structure of interactive concatenative synthesis

The challenge is how we can express the concept of an extremely complex and highly culture-dependent concept of hyperobjects through the synthesis of the sonic idioms. *Concatenative synthesis* is a method of sound synthesis that developed in recent years by the researchers at the Institut de Recherche et Coordination Acoustique/Musique (IRCAM) in Paris, which set the technical foundation for the algorithmic synthesis of sounds described in higher-level terms—terms closer to how we perceive sound, beyond the spectral analysis of audio signals. It is also relatively flexible in terms of defining the units of sound, allowing non-uniform segmentation of the audio stream. The general structure of concatenative synthesis involves: a database of source sound segmented into units, and a unit selection algorithm that finds the units that match the best unit to be synthesized, called the target.[62] (Figure 13)

3.2. Initial Experiments

3.2.1. Interactive Soundscape Composition

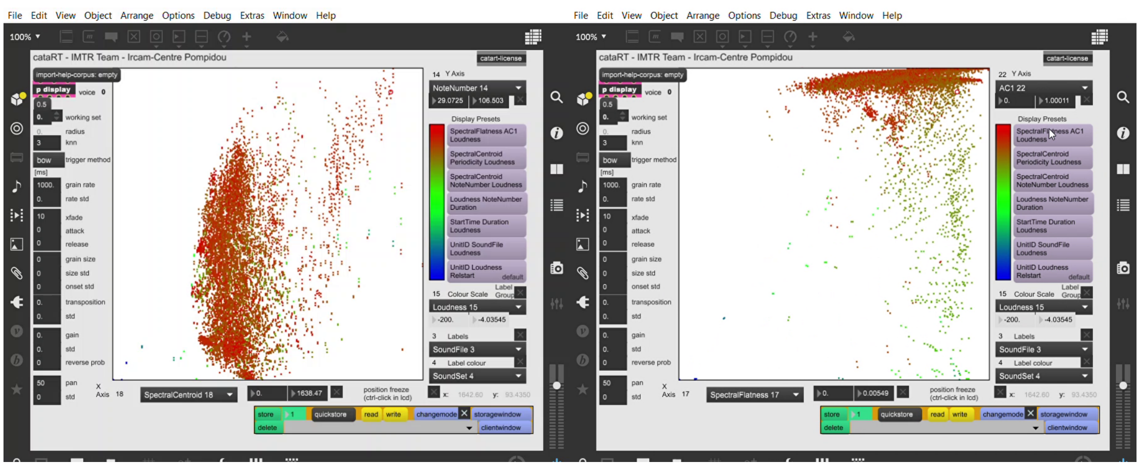


Figure 14. Continuous mapping of 5,000 sound samples using cataRT in Max/MSP

Interactive Soundscape Composition is an experiment to discover the potential of the audio clustering technique and its affordances in creating an imaginary soundscape created from the existing environmental recordings.

---

[62] Schwarz, “Concatenative Sound Synthesis.”

First, I used cataRT, a real-time concatenative synthesis system in Max/MSP developed by Diemo Schwarz.[63] CataRT contains intuitive interfaces and cleverly designed sampling algorithms that allow a variety of schemes to create musical motifs. (Figure 14) The limits of cataRT lie in that it only allows assigning single parameters to the X and Y-axis.

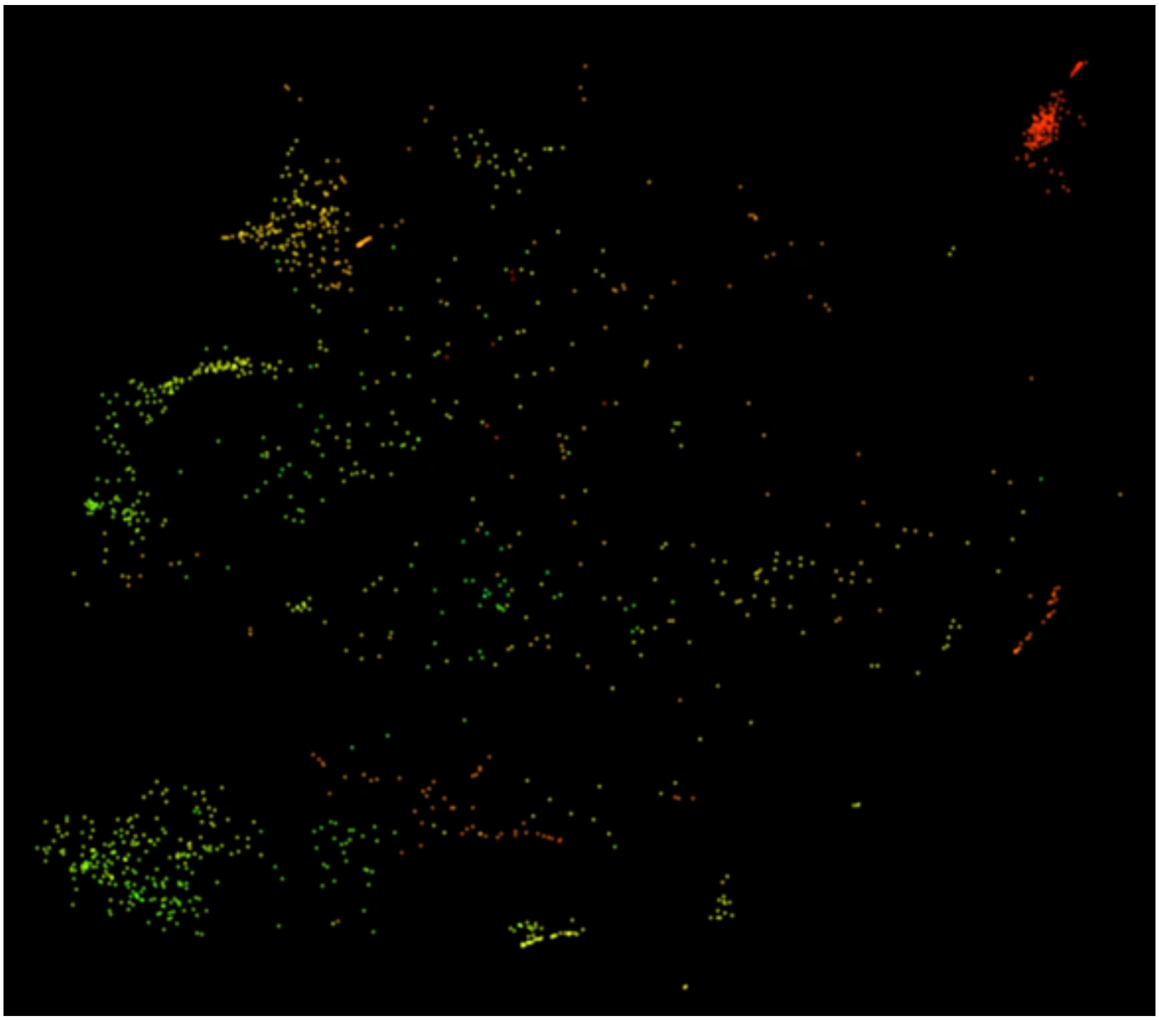


Figure 15. t-SNE plotting of 1,274 audio chunks based on their spectral similarities

---

[63] Diemo Schwarz et al., "Real-Time Corpus-Based Concatenative Synthesis with CataRT," vol. 1 (Internatioanl Conference on Digital Audio Effects (DAFx-06), Montreal, Canada: Oxford University Press, 2006).

I further implemented an interface using Python and JavaScript that allows a more flexible interface in exploring the semantic space of sound. I use *librosa*, a python package for music and audio analysis to extract a feature vector that characterizes each sound clip.[64] Here, the elements of the feature vectors include Mel-Frequency Cepstral Coefficients, first and second-order deltas, and the mean value of each of these across the frames, which will be revisited in 3.3.3. Dimensionality reduction techniques are useful here in mapping the sounds with a feature vector of high dimension onto two-dimensional canvas or three-dimensional space based on their similarity between each other. Figure 15 illustrates 1,274 audio chunks mapped in 2.5 dimensions using T-Distributed Stochastic Neighbor Embedding (t-SNE).

3.2.2. Un-composition

Un-composition is an exploration of the potential of concatenative synthesis technique to discover musical qualities within our everyday environment, or to project the qualities of our everyday sounds to musical instruments. It is both an experiment on *de-composing* the auditory scenes in the environment, and on the dissolved sense of authorship in the process of musical composition. In the context of the thesis, it serves as a prototype to test and evaluate the potentials and limitations of the concatenative synthesis tools.

The experiment was inspired by a Berlin-based Austrian German composer Peter Ablinger's piece *Schaufensterstück* (display-window piece), where he live-feeds the sound captured through a microphone installed at the city center into a mechanically augmented, computer-controlled piano—also known as *Vorsetzer*—as a "display-window."[65]

[64] Brian McFee et al., "librosa: Audio and Music Signal Analysis in Python" (Python in Science Conference, Austin, Texas, 2015), 18–24, https://conference.scipy.org/proceedings/scipy2015/brian_mcfee.html.
[65] Peter Ablinger, "Schaufensterstück," 2004, https://ablinger.mur.at/docu13.html.

Ablinger's piece is largely about the performative spectacle of the contraption playing the piano controlled by the non-human agency, but the musical quality produced throughout the process was limited to MIDI signals playing a virtual instrument. Instead of taking the process of "flattening" the city sounds into MIDI signals that are then transferred to the piano, Un-composition utilizes concatenative synthesis to represent environmental sounds through a variety of acoustic sounds where their textures and timbres remain untouched.

I use AudioGuide, an open-source Python based library for concatenative synthesis developed by Ben Hackbarth, Norbert Schnell, Philippe Esling, and Diemo Schwarz.[66] As described in 3.1.2.2, corpus is the collection of sampled sounds that are then synthesized to match the target, in this case the environmental recordings.

Table 1. Target tracks and their instrumentations in Un-composition

Target track	Instrumentation
Ocean waves	Great bass recorder
Crickets	Flute, violin
Airplane flying overhead	Bass clarinet, bassoon
Train station	Hand percussions (castanets, triangles, woodblocks)
Walking on dry leaves	Hand percussions, viola(pizzicato)

[66] Benjamin Hackbarth et al., "Composing Morphology: Concatenative Synthesis as an Intuitive Medium for Prescribing Sound in Time," *Contemporary Music Review* 32, no. 1 (2013): 49–59.

The instrumentations are then prescribed to each of the tracks based on their perceived similarities between the track and the instrument, and an intended aesthetic purpose (Table 1). For the corpus, I used the Musical Instrument Samples (MIS) created by the members of the Electronic Music Studios at the University of Iowa that has been generously shared since 1997.[67] In addition to MIS, sounds of the great bass recorder, also known as Paetzold recorder, were sampled from the performance of an Italian composer Fausto Romitelli's piece *Seascape*[68], specifically to represent ocean waves.

Then the sites where the target audio is sourced from were carefully selected so that the site's sonic quality is salient enough to be represented through acoustic instruments, including Carlson Beach, Central station in Cambridge, and reservation parks in Boston. The synthesized tracks were then sequenced and mixed to create a "symphony of places".

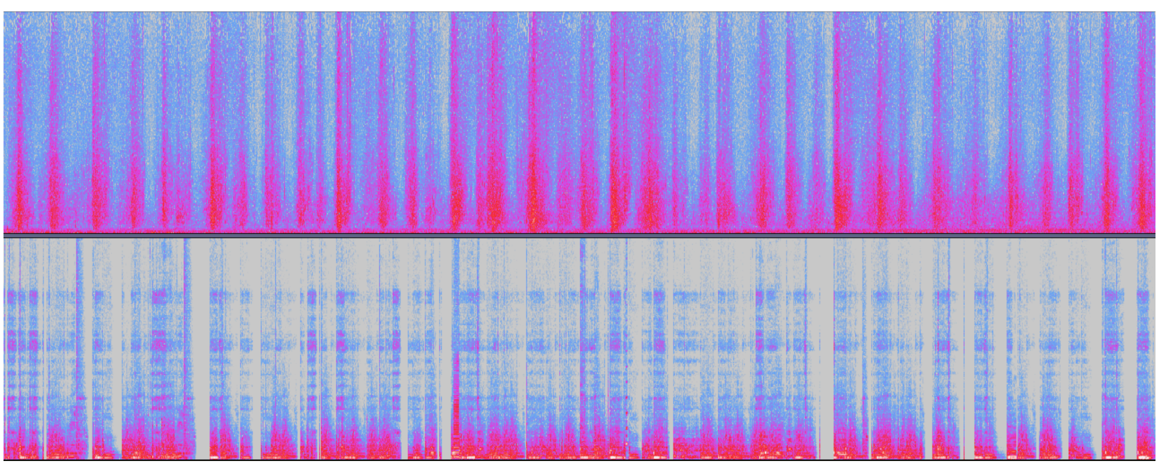


Figure 16. Spectrogram of the ocean wave. Above: original; Below: synthesized

[67] https://theremin.music.uiowa.edu/MIS.html
[68] Ensemble Ictus, "Seascape", 2009, Cypres, track 4 on *Romitelli: Professor Bad Trip*.

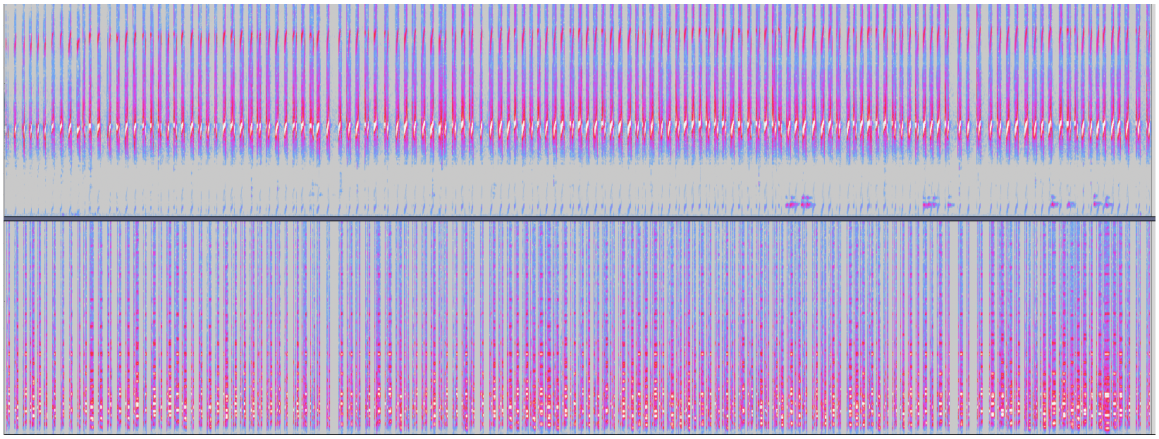


Figure 17. Spectrogram of the crickets. Above: original; Below: synthesized

Among the most interesting results were the ocean waves synthesized from the sounds of the great bass recorder. The instrument exhibits an incredibly complex texture controlled by the blowing techniques of the performer, which resemble the sound of ocean waves – the performer's respiratory air going in and out of the giant wooden box, in sync with the waves coming in and out shore. However, other less periodic and continuous tracks tend to fail to give satisfactory results due to AudioGuide's limitations. The observed limitations throughout the project are as follows.

First, the segmentation conditions of the target audio are very inflexible. The offset of the segments of the target audio is merely defined by two parameters: absolute amplitude threshold and a ratio of the gradient of amplitude. This makes it difficult to process complex sounds from more than one sound source.

Second, while the power of concatenative synthesis lies in the fact that it preserves the performative quality of the original sound, it also compromises the perceptual alignment between the target sound and the synthesized sound. In other words, it is often hard to discern the similarities between the natural and synthesized sounds, especially for complex sounds without clean and discrete fundamental frequencies.

For example, the experiment using the sound of the train approaching the platform as the target audio proved to be very ineffective due to AudioGuide's lack of flexibility in modifying the match. For the instrumental sound to match the gradual crescendo of the train approaching the platform, the corpus should have had audio that was performed in crescendo, which means that in order to get good results for complex and dynamic sounds, an entire collection of audios with different amplitudes, envelopes, played with a whole variety of techniques, is needed.

Therefore, the synthesis involved a few adjustments done manually; first, optional arguments had to be adjusted depending on the type of sound, and second, certain sounds in the corpus had to be manually eliminated. The automated search algorithm often would give unexpected results, which required limiting the sampled corpus. For example, when the corpus representing the ocean waves had flute sounds included along with the recorder, it would match high-pitched flute sound for smaller grains of the segmented target sound, which were prioritized over the recorder due to its arbitrary similarity in its fundamental frequency.

Reflecting on the observations mentioned above, a few of the improvements that should be considered are first, incorporating the capability of modifying the samples while not losing their original performative quality, and second, a more flexible audio segmentation scheme optimized against the alignment between the computation and the human perception.

3.3. Hypermirror: Resonance of the World

3.3.1. Introduction

Hypermirror: Resonance of the World is an interface that allows interactive data probing based on vocal queries. It is a mirror as it responds through a generated audio response based on the audio input—it is a Hypermirror as the "mirror image" is shaped and distorted by the audio database, and by the aural memory of the piece. Data probing is to present a fragment of the data through interaction rather than flattening it into the "god's eye view". It is an experiment to build a sensorium of data and propose a spatial interface to talk to the data through embodied interaction. Again, the concept of *attunement* is useful here—attunement as an act of searching for the frequency that resonates with a certain cluster of data by actively performing your voice.

It is an extension of the concept of Alvin Lucier's work but scaled up beyond physical space. *What are the resonance frequencies of the world?* While following the general process of repeated recording - playing - re-recording to reveal the resonance frequency of the space, certain sounds from the selected audio data slip into the process, creating a sense of expanded resonance and spatiality. The expected result of the system is the sounds from both the audience and the selected audio from the data stacking up incrementally, creating an experience that resembles the "banging-pots-and-pans" example.

3.3.2. System Overview

The general principles of the system design are as follows:

1. **Augment not replace:** There lies an existing rich ecology comprised of software interconnected to one another. The system can include the existing user communities by extending instead of replacing existing software frameworks.

2. **Distribute**: a variety of programming languages have their nature that performs the best for specific purposes, although the boundaries between them are increasingly blurry. For example, processing and synthesizing audio in real-time requires extremely high computational resources in JavaScript or Python.

3. **Hack**: though might be in opposition to the second principle, rethinking the intrinsic limitations of the existing frameworks is the contribution of this implementation. Hacker culture is about contributing to the software ecology by making new connections to the missing links.

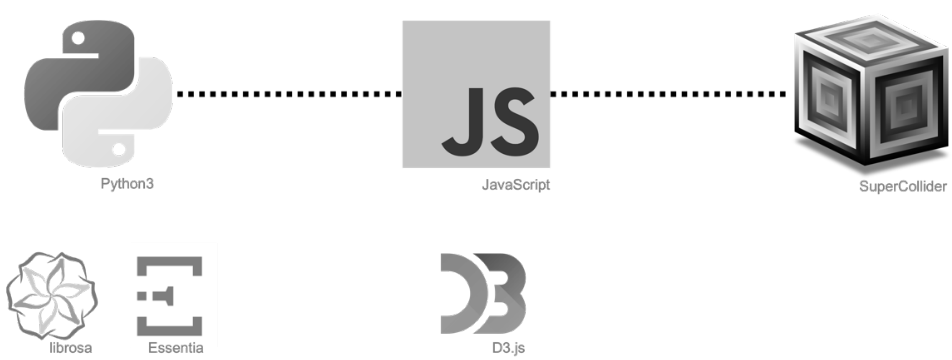


Figure 18. Languages and their libraries used in the system

The overall architecture of computation can be divided into two parts: synchronous and asynchronous. The onset frames within the audio stream are pre-computed with a Python script, which then are passed to the JavaScript to create a web-based visual interface. The web interface then talks to the SuperCollider environment via networking protocol like Open Sound Control (OSC). (Figure 18) The pre-analysis or the representation learning of the audio data takes a significant amount of time that requires it to be asynchronous, but in the future, the system could be integrated into a single, real-time workflow that will be discussed later.

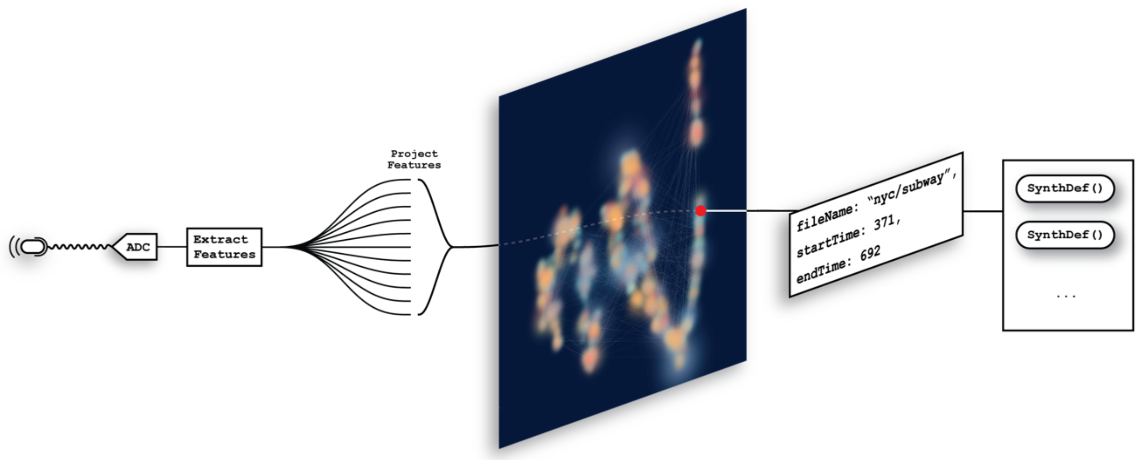


Figure 19. Data flow of sonic mirror

Figure 19 illustrates the data flow of the system. Detailed descriptions of the flow are as follows:

1. The microphone captures the analog signal of the voice input,
2. Analog-to-Digital Converter (ADC) converts the analog input signal to digital,
3. the audio data is processed and analyzed,
4. the feature vector extracted from step 3 is reduced to a point on a graphical interface using the previously learned features,
5. the neighboring points in the existing scatterplot are sampled,
6. the metadata of the points including the file index and timestamp are sent to SuperCollider via OSC,
7. "synthesizers" defined in the SuperCollider script are triggered, generating audio based on the audio database.

3.3.3. Preparing Sounds

3.3.3.1. Onset/Offset Detection

Automatic audio segmentation aims to automate the process of extracting sonic idioms from audio of any length. Onset points marks the start of a new idiom within the audio stream. There are various ways to define onset depending on their purposes. Onset occurs when there is a sudden increase in the prescribed parameter. Essentia offers four onset detection algorithms described in the following:

1. **High Frequency Content (HFC)**: detects changes in the amount of high-frequency content in the audio.[69]
2. **Complex-Domain (CD)**: detects changes in magnitude and phase.[70]
3. **Spectral Flux (SF)**: detects changes in magnitude spectrum.[71]
4. **Magnitude Root Mean Squared (MRMS)**: detects the change in half-rectified RMS of the magnitude.

The four algorithms are designed to serve different purposes: HFC and MRMS detects percussive events, CD takes into consideration both beat and phase by weighting the phase change by its loudness, and SF is more of a general-purpose algorithm. For the purpose of incorporating a wide range of field recordings, I use both CD and SF equally weighted to extract onset frames to generate sonic idioms.

[69] Paul Brossier, Juan Bello, and Mark Plumbley, "Real-Time Temporal Segmentation of Note Objects in Music Signals," 2004.
[70] Juan P. Bello et al., "On the Use of Phase and Energy for Musical Onset Detection in the Complex Domain," *IEEE Signal Rocessing Letters*, 2004, 553–56.
[71] Simon Dixon, "Onset Detection Revisited" (9th International Conference on Digital Audio Effects, Montreal, Canada, 2006), 6.

3.3.3.2. Audio Features

Audio features are the descriptors extracted through the analysis of audio signals. There is a myriad of ways to describe sound computationally, but for the scope of the thesis, I focus on the features that describe timbre to aim for analyzing environmental sounds that do not exhibit salient harmonics and pitch. Timbre, also known as the color of the sound, is a tricky concept to define. It is defined rather in a negative term against the features that can be defined clearly and concisely, such as pitch or loudness.

The perception of timbre is determined by signal spectrum and envelope, which are computationally defined in various ways including spectral features: spectral centroid, skewness, flatness, and rolloff to name a few; harmonic features: inharmonicity and noisiness; and perceptual features: Mel-frequency cepstrum (MFC) that approximates the human auditory system.[72]

Okuyucu, Sert, and Yazıcı identified that ASFCS–H (Spectral Flatness, Centroid, Spread, and Harmonicity) best fits for the purpose of analyzing environmental sounds.[73] While the effectiveness varies depending on the type of sounds that are analyzed, spectral flatness and the MFCC can be used interchangeably.

Note that there is no silver bullet that can fully incorporate the wide range of sounds, and even if there exists a universal solution, it will widely vary depending on the listeners.

---

[72] Geoffroy Peeters et al., "The Timbre Toolbox: Extracting Audio Descriptors from Musical Signals," *The Journal of the Acoustical Society of America* 130, no. 5 (November 2011): 2902–16.
[73] Çigdem Okuyucu, Mustafa Sert, and Adnan Yazici, "Audio Feature and Classifier Analysis for Efficient Recognition of Environmental Sounds," in *2013 IEEE International Symposium on Multimedia*, 2013, 125–32.

3.3.3.3. Dimensionality Reduction

Computational techniques allow us to represent the sounds in a way that is neither symbolic nor spectral (or both symbolic and spectral). Audio data occupies an extremely high dimensional space, and how to reduce it to the dimension directly perceivable to humans has been an important question among the machine learning and the data science community. Dimensionality Reduction (DR) is a process to create a convincing abstraction of the data. We can also speculate DR as the process of creating an "interface" between high-dimensional feature space and perception as it projects the non-perceivable high dimension to the dimension that is directly understandable to humans.

First, the most commonly used, and the oldest DR scheme is Principal Components Analysis (PCA). Its purpose is to compute a set of data with multiple variables and to express it through a set of new, usually two or three variables. These new set of orthogonal variables are called "principal components."[74]

Second, T-Distributed Stochastic Neighbor Embedding (t-SNE) is a DR technique that has been increasingly used since its development in 2008.[75] Unlike PCA, in t-SNE the algorithm deploys a non-linear process where it preserves the local structure of the data, which makes it better for the purpose of clustering.

Finally, Uniform Manifold Approximation and Projection (UMAP) is the newest manifold learning technique. While it is similar to t-SNE but does not rely on the hyperparameters that have to be manually tuned in t-SNE, and also tend to preserve the distances between the clusters better.[76]

---

[74] Hervé Abdi and Lynne J. Williams, "Principal Component Analysis," *WIREs Computational Statistics* 2, no. 4 (2010): 433–59.
[75] Laurens van der Maaten and Geoffrey Hinton, "Visualizing High-Dimensional Data Using t-SNE," *Journal of Machine Learning Research: JMLR.* 9, no. nov (2008): 2579–2605.
[76] Leland McInnes, John Healy, and James Melville, "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction," *ArXiv:1802.03426 [Cs, Stat]*, September 17, 2020.

For the purpose of computing the "mirror image", DR techniques are useful to effectively project the input to search its neighboring points within the database. The matching process involves adding a datapoint to the existing plot, which could be tricky for some DR schemes. Compared to PCA, t-SNE and UMAP are non-linear: they do not yield a function $f(x): R^n \rightarrow R^2$, the function that maps the n dimensional feature vectors into two dimensions. But instead, t-SNE and UMAP let the points "interact" with each other, carrying out different transformations on different regions. Therefore, while adding a new data point in PCA can be done simply by doing simple matrix multiplication, it is a bit more complicated in t-SNE and UMAP.

Figure 20 is an experiment where I segment and cluster the field recording of birds using the DR methods mentioned above. The line connects the audio chunks in chronological order, which graphically represents the overall semantic itinerary within the audio stream.

UMAP preserves global structure better than t-SNE; in other words, the data points far away in the original feature space will remain far away in the reduced representation. While t-SNE is very effective in preserving neighboring correlations, it can also be misleading and unreliable, as Wattenberg, Viégas, and Jonson observed due to its high sensitivity to its hyperparameters, which sometimes have to be manually tuned.[77] Considering the well-tuned balance between the global and local structure of UMAP, it is favorable to other DR schemes in this implementation.

[77] Martin Wattenberg, Fernanda Viégas, and Ian Johnson, "How to Use T-SNE Effectively," *Distill* 1, no. 10 (October 13, 2016): e2.

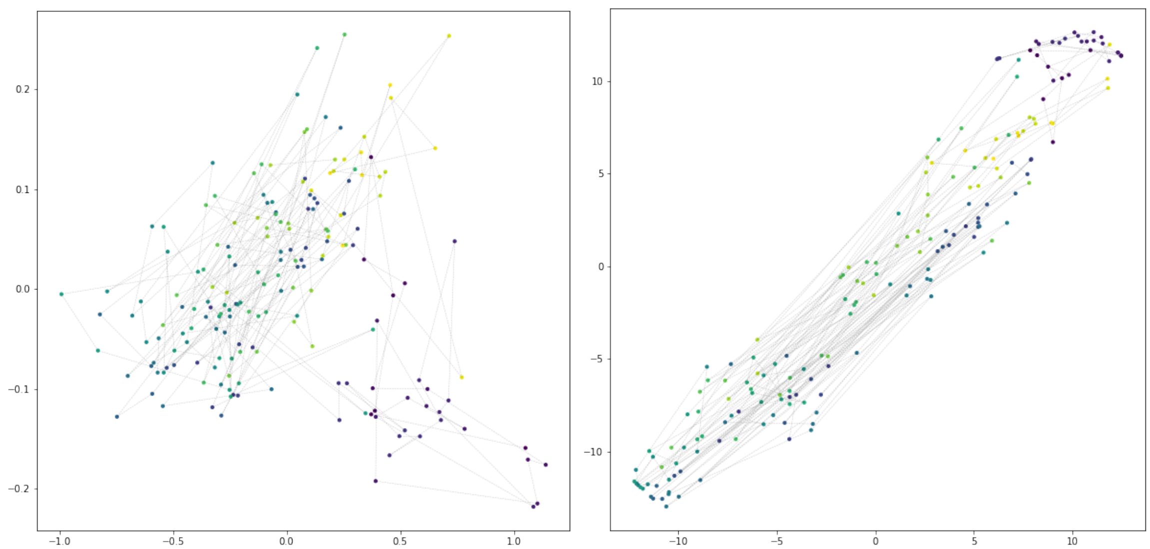
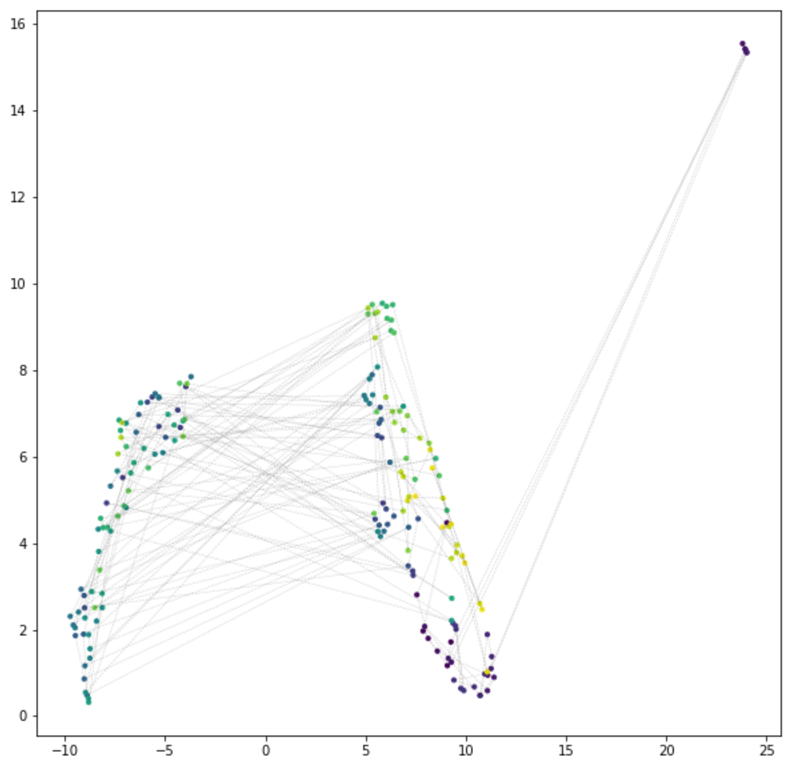


Figure 20. Audio clustering test: PCA (left), t-SNE (right), UMAP (below)



3.3.4. Web Interface Design: Sonic Nebula

Figure 22 illustrates the overall layout of the web interface. It is powered by D3.js for its efficiency, flexibility, and transparency in manipulating Document Object Model (DOM) elements dynamically and interactively within web environment.[78] The interface draws from the motif of nebula—I call it *Sonic Nebula*—which represents the scatterplot's probabilistic nature, which lies first in the process of feature learning and second in the subjectivity of listening. It also allows more overlaps among the datapoints for smoother interpolation between them.

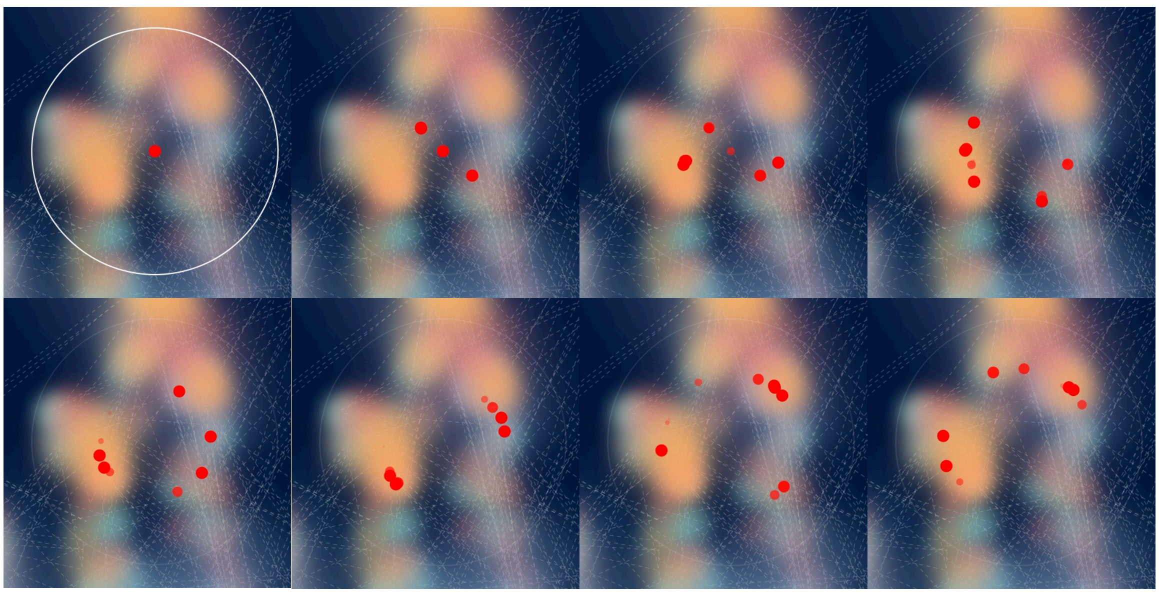


Figure 21. Series of screenshots of web interface upon receiving the voice input

The point ripples throughout the existing plot upon receiving the voice input, which is located at the coordinate computed through the learned feature embedding matrix. Sampled points are marked red and diminish in their size and opacity on the prescribed rate, along with the decaying audio (Figure 21).

---

[78] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer, "D³ Data-Driven Documents," *IEEE Transactions on Visualization and Computer Graphics* 17, no. 12 (December 2011): 2301–9.

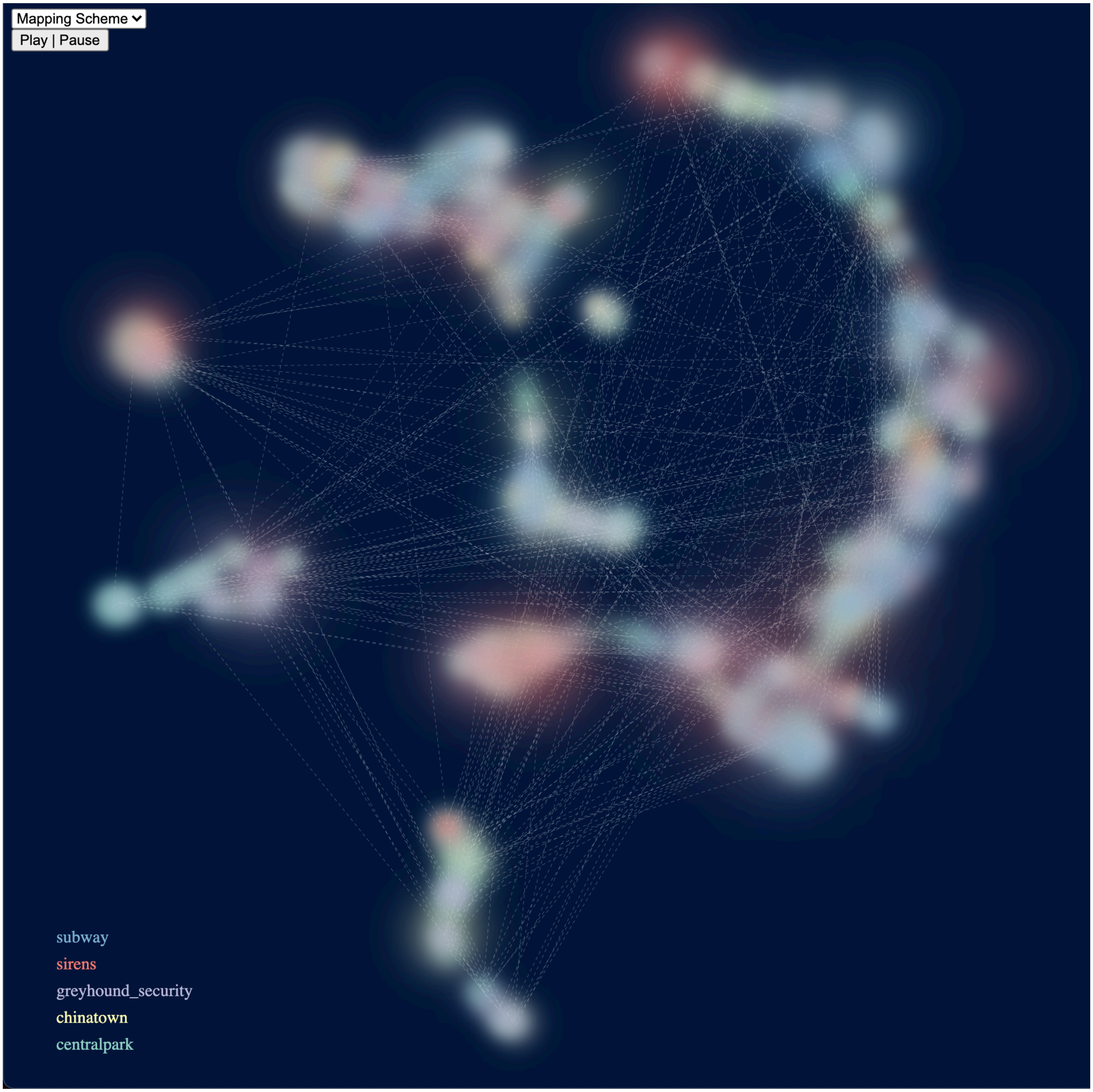


Figure 22. Overview of the web interface

The points are color-coded by the source of the audio. A dropdown menu is added to allow the users to choose among the DR schemes to help them understand how each option behaves differently.

3.3.5. Experiments

Following are the experiments utilizing the developed interface. Audio recordings were collected based on the curated themes. Longer audios were edited down to less than two minutes to prevent excessive number of duplicate segments.

3.3.5.1. Aquasphere

First experiment is entitled *Aquasphere*, where I collected a range of sounds that are in any way related to water, including river streams, raindrops, and glaciers melting. (Table 2) Note that the average number of segments per minute is much larger for the sound of the raindrops than the river as the audio signature of raindrops is more staccato-like, whereas it is more continuous for the river.

Table 2. List of audio in Aquasphere

Description	**Geotag**	**Duration**	**Segments**	**Source**
Small stream in a forest	Mátrafüred, Hungary	0:24	29	freesound.org/s/509174/
River with a strong stream	Ustou, France	2:00	175	freesound.org/s/243487/
Raindrops on the rooftop	Berlin, Germany	0:37	281	freesound.org/s/204473/
Glacier melting	Fláajökull, Iceland	0:28	122	audioboom.com/posts/5176415
	Total	3:29	607	

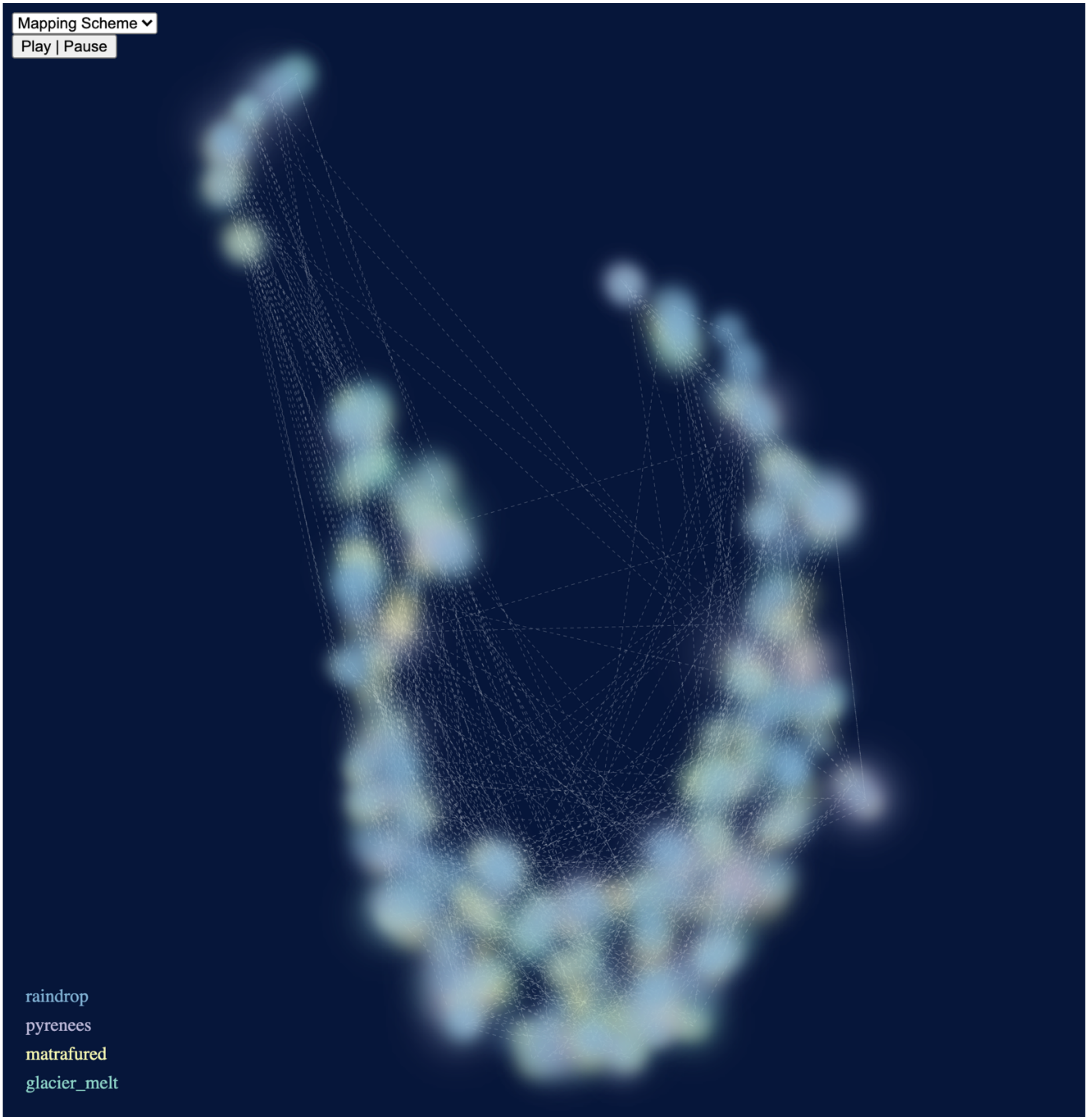


Figure 23. Aquasphere in UMAP view

3.3.5.2. New York City Sonosphere

New York City Sonosphere is an experiment to create a sonic representation of a metropolitan city. Compared to *Aquasphere*, the collection process was constrained based on the geographical region, in this case New York City. I wanted to encapsulate a full color range of the lives in the city, and the recordings include musical performances in Central Park and Chinatown, subway trains, and sirens.

Table 3. List of audio in New York City Sonosphere

Description	**Geotag**	**Duration**	**Segments**	**Source**
Crowd playing instruments in Central Park	Central Park	1:36	326	audioboom.com/posts/7435493
Chinese string instrument performance in Columbus Park	Columbus Park, Chinatown	0:45	82	audioboom.com/posts/7432425
Greyhound bus security announcement	(New York City)	1:22	159	freesound.org/s/203112/
Firetruck sirens	(New York City)	1:35	43	audioboom.com/posts/7427209
Subway journey	110th St. - 138th St.	0:59	82	audioboom.com/posts/3151247
	Total	6:17	692	

3.3.5.3. Results

The most exciting moments were when multiple sounds from different sources were combined, transporting the listeners into the virtual space where different times and spaces were merged.

In the case of NYC Sonosphere, the listeners were able to travel around the different facets of lives in NYC in a fragmented yet wholistically. There were also no distinctions between the data and the creators of the data, the people and the environment that contributed to the sonic environment in which the recordist was situated.

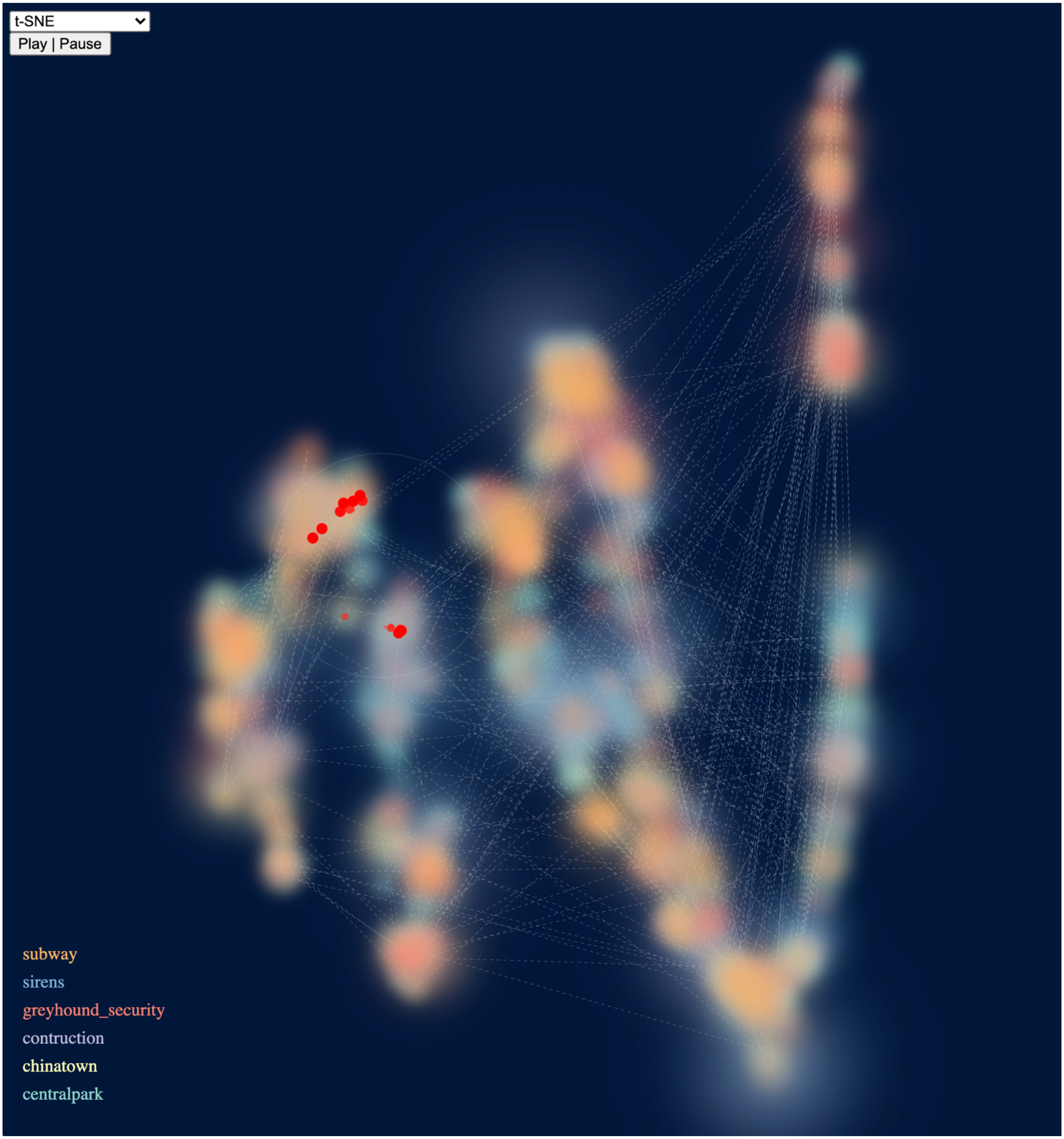


Figure 24. New York City Sonosphere in t-SNE view

04 Discussion

4.1. Potential Usage

4.1.1. Sound Designer's Toolbox

Sonic Hypermirror can be utilized as a useful tool for the practice of sound design. The most common example of sound design is foley in filmmaking, the art and practice of creating and mixing a convincing sound that matches the movie scene.

The process of foley production involves messy experiments tapping, rubbing, and banging hundreds of objects against multiple objects or surfaces in order to choose the most plausible sound among them. To boil down hundreds of options and make a final judgment is extremely tricky. Using the Sonic Nebula interface, sound designers can have their growing archive of recorded sounds from in and out of the studio, where they can explore the best matching sound by mimicking the desired tonal color.

4.1.2. Sonic Cartography

Sonic Nebula can be used for "sonic cartography" of cities. Sonic cartography is mapping the sonic environment into a representation that is in between a geographical map and the sound contained within each geographical region. As seen in the *New York City Sonosphere*, sound visualization serves as an interface to portray sonic events and activities. By manipulating the prescribed region, you can start to compare the aural quality between the regions.

4.2. Future Work

4.2.1. Integration

A few open-source, web-based sound processing libraries are recently developing in progress. For example, Essentia.js developed by members of the MTG, offers a high-level API for signal processing and music information retrieval within the web environment.[79] They will potentially allow adding new audio or even live streaming into an existing database in real-time and integrating the I/O in the web-based platform.

4.2.2. Going Even Bigger

The community of environmental sound recording has been growing fast in recent years, and there are several online platforms that are expanding rapidly in

[79] Albin Correya et al., "Audio and Music Analysis on the Web Using Essentia.Js," *Transactions of the International Society for Music Information Retrieval* 4, no. 1 (November 22, 2021): 167–81.

their sizes. The platforms include Cities and Memory[80], Freesound[81], Sound Around You to mention a few.

The one with the biggest repository so far is Freesound, an open-source, collaborative library run by Music Technology Group (MTG) at Universitat Pompeu Fabra in Barcelona. Freesound has more than 500,000 and growing audio recordings and sound effects. Its API offers data querying of the database based on the audio's metadata that includes tags and some of them geotags added by users, which will enable parsing of a large amount of audio data grouped by its content and its geographical locations.

4.2.3. Revisiting Amacher's City-Links

Amacher describes her work City-Links as "a collection where sounding resources of two or more remote locations (cities or locations within a city) are fed back to each other to allow for interaction between men and sound at distant locations."[82]

I imagine where an array of microphones is installed at distributed locations in two selected cities, and the recordings are fed into the audio data in real-time. Using the exact same system but by replacing the voice with sounds of a city, it can create a feedback link between the two cities mirroring each other, gauging the similarities and differences between the acoustic environment of distant locations.

[80] Founded by sound artist Stuart Fowkes in 2014 specifically to share field recordings usually longer than a minute. It has more than 5,000 sounds from more than 1,000 contributors as of 2021. https://citiesandmemory.com.
[81] https://freesound.org/
[82] Maryanne Amacher, "Long Distance Music," 105, in Dietz and Cimini, *Maryanne Amacher: Selected Writings and Interviews*, 94.

05 Conclusion

In the flood of images, we are rather becoming more and more blind. In the thesis, I reiterate Donna Haraway's criticism of data visualization as "god trick," and foreground audition as a primary sensory modality hoping to overcome the "partial perspective" and to offer the "close-to-god" experience. Some things are better heard than seen.

In the context of the motivation stated above, I presented a novel interface Sonic Hypermirror, a scalable data-probing tool using vocal query. The interface brings a large amount of data into a single sphere of perception. It does not give a wholistic "view" of the data; it rather urges the audience to engage and explore the data, which I call "attunement." A variety of techniques to analyze and process field recordings were presented; with them I defined the concept of "sonic idiom," the phrase for building a poetry of sound, aiming to bring linguistic ideas to sound composition.

It is followed by an extensive theoretical background synthesizing the vast amount of literature around politics of data, architecture, sound studies, computer music, and human-computer interaction. I tackle the issue of hyperobject coined by Timothy Morton, how vision has failed to deliver the high-dimensional phenomena and how sound is better in delivering complex, emotionally charged ideas. Audition is proven more effective in evoking empathy in a way that is softer and more rhetorical than the "transplantation" of the entire experience.

The contributions of the work are threefold: I proposed the system of data representation that is continuous, non-referential, and exploratory; revisited the affordances of architectural space as a data storage and an interactive datascape; and presented the novel interface that allows embodied interaction with the scale that normally cannot be experienced directly.

06 List of Figures

07 Bibliography

Abdi, Hervé, and Lynne J. Williams. "Principal Component Analysis." *WIREs Computational Statistics* 2, no. 4 (2010): 433–59.

Ablinger, Peter. "Schaufensterstück," 2004. https://ablinger.mur.at/docu13.html.

Arte TRACKS. *SOPHIE: The Producer Taking Pop to the Future (English Version / Interview) | Arte TRACKS*, 2018. https://www.youtube.com/watch?v=2ifh0tDrwBA.

Ball, Thomas S., and L. C. Bernardoni. "Application of an Auditory Apperception Test to Clinical Diagnosis." *Journal of Clinical Psychology* 9 (January 1953): 54–58.

Barthes, Roland. "Listening." In *The Responsibility of Forms: Critical Essays on Music, Art, and Representation*, 1st ed., 245–60. New York: Hill and Wang, 1985.

Beard, David. "Acousmatic Listening and a Critical Awareness of Place." *International Journal of Listening* 33, no. 3 (September 2019): 129–32.

Bello, Juan P., Chris Duxbury, Mike Davies, Mark S, and Senior Member. "On the Use of Phase and Energy for Musical Onset Detection in the Complex Domain." *IEEE Signal Rocessing Letters*, 2004, 553–56.

Bostock, Michael, Vadim Ogievetsky, and Jeffrey Heer. "D^3 Data-Driven Documents." *IEEE Transactions on Visualization and Computer Graphics* 17, no. 12 (December 2011): 2301–9.

Braidotti, Rosi. "'We' Are In This Together, But We Are Not One and the Same." *Journal of Bioethical Inquiry* 17, no. 4 (December 2020): 465–69.

Braverman, Sydell, and Hector Chevigny. *The Braverman-Chevigny Auditory Projective Test*. American Foundation for the Blind, Inc., 1951.

Bregman, Albert S. *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge: The MIT Press, c1990.

Brossier, Paul, Juan Bello, and Mark Plumbley. "Real-Time Temporal Segmentation of Note Objects in Music Signals," 2004.

Chion, Michel. "The Three Listening Modes." In *Audio-Vision : Sound on Screen*, 22–34. New York: Columbia University Press, 1994.

Coleman, Nancy. "On the Front Page, a Wall of Grief." *The New York Times*, February 21, 2021, sec. Times Insider. https://www.nytimes.com/2021/02/21/insider/covid-500k-front-page.html.

Correya, Albin, Jorge Marcos-Fernández, Luis Joglar-Ongay, Pablo Alonso-Jiménez, Xavier Serra, and Dmitry Bogdanov. "Audio and Music Analysis on the Web Using Essentia.Js." *Transactions of the International Society for Music Information Retrieval* 4, no. 1 (November 22, 2021): 167–81.

Cowen, Alan, Hillary Elfenbein, Petri Laukka, and Dacher Keltner. "Mapping 24 Emotions Conveyed by Brief Human Vocalization." *American Psychologist*, December 20, 2018.

Davison, Paul, and Rohan Seth. "Clubhouse — Check 1, 2, 3... Is This Thing On?" Clubhouse. Accessed April 23, 2021. https://www.joinclubhouse.com/check-1-2-3.

Delalande, François. "Les Unités Sémiotiques Temporelles: Problématique et Essai de Definition." In *Les Unités Sémiotiques Temporelles*. Marseille: MIM, 1996.

Dietz, Bill, and Amy Cimini. *Maryanne Amacher: Selected Writings and Interviews*. Brooklyn, NY: Blank Forms Editions, 2021.

Dixon, Simon. "Onset Detection Revisited," 6. Montreal, Canada, 2006.

"DRAM: Notes for 'Alvin Lucier: I Am Sitting in a Room.'" Accessed January 20, 2022. https://www.dramonline.org/albums/alvin-lucier-i-am-sitting-in-a-room/notes.

Fauconnier, Gilles, and Mark Turner. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. 1. paperback ed. New York, NY: Basic Books, 2002.

Foucault, Michel. "Of Other Spaces." Translated by Jay Miskowiec. *Diacritics* 16, no. 1 (1986): 22–27.

Gabor, D. "Acoustical Quanta and the Theory of Hearing." *Nature* 159, no. 4044 (May 1947): 591–94.

Gaver, William. "Auditory Icons: Using Sound in Computer Interfaces." *Human-Computer Interaction* 2, no. 2 (June 1, 1986): 167–77.

Gaver, William W. "What in the World Do We Hear? An Ecological Approach to Auditory Event Perception." *Ecological Psychology* 5, no. 1 (1993): 19.

Gaver, William W., Jacob Beaver, and Steve Benford. "Ambiguity as a Resource for Design." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 233–40. CHI '03. New York, NY, USA: Association for Computing Machinery, 2003.

Goguen, Joseph. "An Introduction to Algebraic Semiotics, with Application to User Interface Design." In *Computation for Metaphors, Analogy, and Agents*, edited by Chrystopher L. Nehaniv, 1562:242–91. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999.

Hackbarth, Benjamin, Norbert Schnell, Philippe Esling, and Diemo Schwarz. "Composing Morphology: Concatenative Synthesis as an Intuitive Medium for Prescribing Sound in Time." *Contemporary Music Review* 32, no. 1 (February 1, 2013): 49–59.

Haraway, Donna. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." *Feminist Studies* 14, no. 3 (1988): 575.

Hogan, Robert. "Development of an Empathy Scale." *Journal of Consulting and Clinical Psychology* 33, no. 3 (1969): 307–16.

Holmes, Thom. *Electronic and Experimental Music: Technology, Music, and Culture*. 3rd ed. New York: Routledge, 2008.

Johnson, Mark. *The Meaning of the Body: Aesthetics of Human Understanding*. Chicago, IL: University of Chicago Press, 2008.

Katz, Mark. "Music in 0s and 1s: The Art and Politics of Digital Sampling." In *Capturing Sound: How Technology Has Changed Music*. University of California Press, 2004.

Ken Murphy. *A History of the Sky*, 2011. https://www.youtube.com/watch?v=PNln_me-XjI.

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

Lakoff, George, and Mark Johnson. *Metaphors We Live By*. Chicago: University of Chicago Press, 1980.

Maaten, Laurens van der, and Geoffrey Hinton. "Visualizing High-Dimensional Data Using t-SNE." *Journal of Machine Learning Research : JMLR.* 9, no. nov (2008): 2579–2605.

McCutchon, Tommy. "Unlimited Americana: A Conversation with Halim El-Dabh." Music & Literature, June 2017. https://www.musicandliterature.org/features/2017/6/1/unlimited-americana-a-conversation-with-halim-el-dabh.

McFee, Brian, Colin Raffel, Dawen Liang, Daniel Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto. "Librosa: Audio and Music Signal Analysis in Python," 18–24. Austin, Texas, 2015.

McInnes, Leland, John Healy, and James Melville. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction." *ArXiv:1802.03426 [Cs, Stat]*, September 17, 2020. http://arxiv.org/abs/1802.03426.

Mellen, Ruby. "How Pots and Pans Became Tools of Protests, from Chile to Myanmar." *Washington Post*, February 4, 2021. https://www.washingtonpost.com/world/2021/02/04/pots-pans-protests-myanmar-coup/.

Milk, Chris. *How Virtual Reality Can Create the Ultimate Empathy Machine*, 2015. https://www.ted.com/talks/chris_milk_how_virtual_reality_can_create_the_ultimate_empathy_machine.

Mills, Mara. "Evocative Object: Auditory Inkblot." *Continent* 5, no. 1 (2016): 15–23.

Morton, Timothy. "Attunement." *Ecology Without Nature* (blog), November 2, 2014. http://ecologywithoutnature.blogspot.com/2014/11/attunement.html.

———. *Hyperobjects: Philosophy and Ecology after the End of the World*. University of Minnesota Press, 2013.

Obadike, Mendi, and Keith Obadike. *Numbers Station 1 [Furtive Movements]*, 2016. https://www.youtube.com/watch?v=PuLzv53gM_o.

Okuyucu, Çigdem, Mustafa Sert, and Adnan Yazici. "Audio Feature and Classifier Analysis for Efficient Recognition of Environmental Sounds." In *2013 IEEE International Symposium on Multimedia*, 125–32, 2013.

Oliveros, Pauline. "Quantum Listening: From Practice to Theory (To Practice Practice)," 2000.

Peeters, Geoffroy, Bruno L. Giordano, Patrick Susini, Nicolas Misdariis, and Stephen McAdams. "The Timbre Toolbox: Extracting Audio Descriptors from Musical Signals." *The Journal of the Acoustical Society of America* 130, no. 5 (November 2011): 2902–16.

Reydellet, Jean de. "Pierre Schaeffer, 1910-1995: The Founder of 'Musique Concrète.'" *Computer Music Journal* 20, no. 2 (1996): 10–11.

Rodgers, Tara. "Approaching Sound." In *The Routledge Companion to Media Studies and Digital Humanities*, edited by Jentery Sayers, 1st ed., 233–42. New York : Routledge, Taylor & Francis Group, 2018.: Routledge, 2018.

Roger Ebert on Empathy. Accessed January 12, 2022. https://www.rogerebert.com/empathy/video-roger-ebert-on-empathy.

Schwarz, Diemo. "Concatenative Sound Synthesis: The Early Years." *Journal of New Music Research* 35 (March 1, 2006): 3–22.

Schwarz, Diemo, Gregory Beller, Bruno Verbrugghe, and Sam Britton. "Real-Time Corpus-Based Concatenative Synthesis with CataRT," Vol. 1. Montreal, Canada: Oxford University Press, 2006.

Seachrist, Denise A., Halim El-Dabh, and George Faddoul. *The Musical World of Halim El-Dabh.* Kent, Ohio: Kent State University Press, c2003., 2003.

Shalabi, Sam, and Halim el-Dabh. "Step Into the Electric Magnet: Music Pioneer Halim El-Dabh." *Bidoun: Arts and Culture from the Middle East* 1 / 9 (January 1, 2007): 78–80.

Silvans S. Tomkins. *The Thematic Apperception Test: The Theory And Technique Of Interpretation*, 1947. http://archive.org/details/in.ernet.dli.2015.60536.

Sterne, Jonathan. "Sonic Imaginations." In *The Sound Studies Reader*, 1–17. New York: Routledge, 2012.

———. "The Stereophonic Spaces of Soundscape." In *Living Stereo: Histories and Cultures of Multichannel Sound*, 65–83. New York: Bloomsbury Academic, 2015.

Strachey, Sir Richard. "The Air Waves and Sounds Caused by the Eruption of Krakatoa in August, 1883." In *The Eruption of Krakatoa: And Subsequent Phenomena*, 57–88. Trübner & Company, 1888.

TEDx Talks. *The Difference between Hearing and Listening | Pauline Oliveros | TEDxIndianapolis*, 2015. https://www.youtube.com/watch?v=_QHfOuRrJB8.

The Pulsa Group. "The City as an Artwork." In *Arts of the Environment*, 1972. http://archive.org/details/TheCityAsAnArtwork.

Turkle, Sherry. *Evocative Objects: Things We Think With*. Cambridge, Mass: MIT Press, 2007.

Voegelin, Salomé. *Sonic Possible Worlds: Hearing the Continuum of Sound*. New York London New Delhi Sydney: Bloomsbury, 2014.

Wakin, Daniel J. "'Those We've Lost,' a Chronicle of Covid Death, Comes to a Halt." *The New York Times*, June 4, 2021, sec. Times Insider. https://www.nytimes.com/2021/06/04/insider/covid-obituaries-those-lost.html.

Walkin, Daniel J. "Those We've Lost." *The New York Times*, April 16, 2020, sec. Obituaries. https://www.nytimes.com/interactive/2020/obituaries/people-died-coronavirus-obituaries.html.

Wattenberg, Martin, Fernanda Viégas, and Ian Johnson. "How to Use T-SNE Effectively." *Distill* 1, no. 10 (October 13, 2016): e2.

Zer-Aviv, Mushon. "DataViz—The UnEmpathetic Art." *Responsible Data* (blog), October 19, 2015. https://responsibledata.io/2015/10/19/dataviz-the-unempathetic-art/.