Revealing Network Infrastructure at Geographic Scale Using Location Based Audio

John Brumley

University of Tsukuba
Tsukuba, Japan
john@johnbrumley.info

Abstract

Data centers are a necessary element of contemporary global network infrastructure, but are generally overlooked due to obscurity or more often indifference by the general public. This project aims to elevate the importance of data centers within urban landscapes by providing them with unique sound signatures. To achieve such a goal, a web based application has been developed that takes a user's location and orientation and creates a synthesized three-dimensional audio space based on that user's spatial relation to nearby data centers. This enables a user to passively listen to an audio representation of the combined virtual activity within a given region. Additionally, users can actively seek out individual data centers using their emanations as a navigation tool.

Keywords

Geolocation, Augmented Reality, Infrastructures, Data Center, Sonic Environment, 3D Audio, Mobile Phone, Web Application

Introduction

Network protocols, which enable computer-to-computer communication and thus mediate everything that travels over a digital network, perform silently and are often highly abstracted and simplified when a user receives a notification. Unless an application is specifically for network traffic analysis, it would be inappropriate for a developer or designer to include such traffic as part of its graphical interface. Yet to achieve a solid awareness of digital space and to develop abilities for traversing and modifying hybrid spaces, one cannot remain oblivious to the underlying structures and semaphore of digital communication.

There exist tools for uncovering the particulars of networking traffic, yet are, for the most part, for system administrators and require some preexisting understanding of network jargon to make much sense of a stream of data (e.g. Wireshark, Carnivore Client, Little Snitch). These types of software are indispensable for monitoring and packet analysis. They can provide indirect reminders that routers and data centers are functioning properly, and can even describe the pathways taken by individual packets via commands like *traceroute*. However, these tools only function when a user explicitly wants to discover network traffic and consequently fail to represent the continuous existence of the infrastructural elements that facilitate such data transfer whether the user considers them or not.

In the domain of auditory displays, most of the previously mentioned tools only provide a minimal amount of sonified information. These types of output are used to assist in monitoring a network and highlight important events such as a network aberration or a remote access request. As such, the sounds are discrete and meant to alert the user of an action that needs to be taken (Walker & Nees, 2011). While successful at conveying that a network event has occurred, the audible components in isolation fail to provide any added understanding or insight regarding the overall networked system or its relationship to outside systems.

Even when successful at depicting the constant flow of data, monitoring software remains primarily within the physical boundaries of screen space. Ignoring the physical nodes and edges supporting our networks presents a detached and incomplete perspective for the user. It is not sufficient to simply provide the positions, addresses, or satellite views of the buildings (data centers) housing this equipment, as it fails to elicit an embodied relationship with infrastructure based on the physical constellation of the collected human and nonhuman actors. In the same way that the stepping points of a network are inert until activated (performed), the experience of the surrounding data centers should also be revealed through the active participation in space with those sites (Thrift, 1997).

By using a portable device, such as a smart phone, to add perceivable characteristics to a data center *entity*, one can sense the presence of data centers and gain a more concrete understanding of their infrastructural surroundings. Venturing into physical space to encounter data centers lets users achieve experiential and embodied understanding of their relationship to infrastructure. Because sound can so readily fade into and out of our focus, yet remain ever present, it provides the most effective characteristic for a data center to virtually emit.

Implementation

Adding audio characteristics to the data centers is achieved through the use of stereo headphones attached to a smart phone capable of running a web-based application. The web application is the primary method for connecting a user's position (location and heading) to an existing data-base of actual data centers.

Figure 1. The PannerNode object takes multiple parameters. This project specifies a position while leaving the sound cone omnidirectional. Image by Mozilla Contributors is licensed under CC-BY-SA 2.5

Location

When a user connects to the application, the server queries an up-to-date repository of publicly listed data centers throughout the world, checking against a database stored on the web app's server for any changes in data center locations. To reduce latency for the end user, the server returns a JSON object with a subset of centers filtered based on the user's location. In order to provide a user's position, the user must grant the application permission to access her device's GPS and orientation sensors. Once the locations of both the user and the data centers have been retrieved, the client side of the application can initialize the audio.

Audio

All audio is generated and controlled using the JavaScript Web Audio API which allows web-based applications to take advantage of much more sophisticated control over audio without the use of plugins.

Following the location data retrieval, an audio context is initialized and the position of the user, a vector containing latitude and longitude coordinates, is assigned to the main *AudioListener* object. Additionally, data collected from device orientation sensors is processed to produce a compass heading ranging from 0 to 360 degrees and assigned to the *AudioListener* orientation. Both of these values are updated whenever new sensor data is made available.

For each data center, a *PannerNode* is created based on its respective global coordinates. The directionality of sound moving away from the node is controlled by the shape of a sound cone (Figure 2), however in this project all nodes are set to be omnidirectional. The node on its own does not emit any audio, therefore once a node is created it will immediately be assigned a unique sound.

With both an *AudioListener* and at least one *PannerNode* created, any audio produced by a node will be filtered and attenuated based on the orientation of the listener and the distance model of the respective node. The *PannerNodes* will remain fixed, while all changes in audio are determined by the listener's change of position and orientation. As the user moves through a region, whether by train, car, or on foot, the levels and positions of the audio will shift and change, creating unique mixes of the various centers within that region.

The head -related transfer function (HRTF) used by the Web Audio API to provide accurate spatial discrimination for the user, using impulse responses from human subjects (Adenot & Toy, 2016), did not create a significant enough difference between sounds located directly in front of a user and sounds located directly behind a user. In systems where small changes in position provide clear indications regarding the distance of a user to a sound object, the problem of inverted direction can be detected and corrected quickly by the user (Carlile, 2011, p. 54). However, when the amplitude falloff of a sound may only become noticeable after hundreds of meters of movement, as in this system, clarity of a sound object's spatial position will be the primary source of information used in navigation. Because of this, further processing of the sound based on the orientation is necessary to prevent misunderstandings and frustration for the system's users.

The decision to have such parameter mapping, to represent extremely slow and gradual auditory feedback for users of the system, is reflective of the geographic size of the region exposed by the auditory display. Given such a slow feedback loop, and provided that the collective data center soundscape maintains an amount of interest, the full sonic capacity of an individual data center can slowly emerge. A visual analog to this experience might be driving toward a benign looking mountain, far along the horizon, and gradually realizing its imposing stature as you move ever closer. Rather than downplay the importance and presence which data centers hold in our lives by using a more immediate audio perception action-loop, their importance is magnified through the energy we must expend in order to significantly modify the sonification model. This highlights the affective nature of spatialized sound by combining the auditory qualities of the sound sources (see below) and the performative nature of such a labor-intensive perception-action loop. However, it is also possible that because the interaction design of this system goes against the transparency between action and effect which characterize most successful sonic interaction designs, that participants may become frustrated or disinterested by lack of immediate feedback (Serafin et al., 2011).

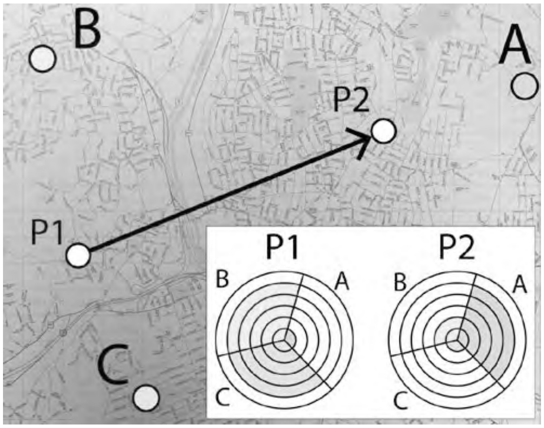

Figure 2. A representation of the change in sound influence of data centers A, B, and C as a listener moves from point P1 to P2. Each center's color gradient shows the size of its sound influence based on a user's distance from that center. The diagrams show approximate loudness of each center for each point

Sound

The sounds attached to each *PannerNode* are created from modified audio samples with adjustments to the buffer speeds of the samples based on the relationship of the listener to the data center and other listeners. Historically, there have been many examples of communication devices being used as tools for extracting sounds, either incidental to the device or indicative of the network's medium. From Thomas Watson listening to natural radio through a telephone wire to works by Paul DeMarinis, e.g. *Rome to Tripoli* (2006), these examples both channel and examine the natural energy of electromagnetism that saturates our atmosphere (Kahn, 2013). With the current dominance of digital communication protocols adding a layer of separation from the analog, it becomes necessary to also explore the underlying components that drive digital communication. Therefore, the methods for generating sounds which intend to explore the energetic activity of contemporary networks, in the same vein as the above examples, should similarly embrace the systems, protocols, and physical components which underpin them. Because the focus of this project is to consider the spatial relationships of physical bodies and structures, the sounds draw from the hard-ware and mechanical qualities of the network rather than exploring the higher-level messaging protocols or soft-ware.

Several years ago, I was given the opportunity to visit two separate data centers on the campus of a university. I was fascinated at the contradicting nature of the space. The visual stillness of the racks of servers, the neatly strung cables, and the uniform fluorescent lighting was immediately overshadowed by the filtered noise of the arrays of fans simultaneously pumping air through the machines. It brought to mind the dichotomy of the static and solid external appearance of the computer against the inner chaos of the CPU, GPU, and hard drive produce billions of operations per second. This relationship of an unremarkable exterior belying chaotic internal activity informed the process of creation for the sounds in this project.

Each sound begins as a simple audible waveform, with a relatively low frequency, recorded into a buffer. The speed of the sample is then multiplied by a factor to bring the frequency of the waveform far above the range of human hearing (20-20,000Hz) and into the same frequency spectrum of computer hardware operation speeds (MHz, GHz). These goal frequencies are determined by various mile-stones within the history of computing systems, for example the original IBM PC had a clock speed of 4.77MHz. The radical shift in frequency, coupled with the degradation of the sample inherent in the process of modifying the sound, results in a drone of subharmonics both chaotic and stable.

Rather than sonifying the user's position and orientation information solely from the perspective of communication efficiency, using beacon chimes or spoken descriptions of the occurrences at each location, the sonic nature of the overall soundscape depicts the continuous activity of a large network ecosystem. This presents problems regarding audio stream segregation when many different data centers are emitting audio in close range to one another. Though spatialization is usually aids in the process of separating sounds (Neuhoff, 2013), the complex and continuous nature of the sounds require further assistance with differentiation of each audio stream. A solution is found in the bioacoustics of rainforests, where the large numbers of vocalizing animals create a highly-crowded frequency spectrum. Rainforest species have adapted to occupy "niche" frequency bands within the overall audio spectrum, which allows them to communicate with others of the same species within their own audio territories (Krause, 2011). Just as different species of animals within a rainforest differentiate their sounds using their own species-specific frequency bands, the different datacenters of a metropolitan region can each inhabit their own bands within the frequency spectrum. Combining virtual sound spatialization with bioacoustics inspired frequency differentiation will aid users in pinpointing each unique sound source while not detracting from the overall auditory scene.

Conclusion

By creating a system that enables data centers to emit virtual sounds across large distances, I hope to allow users to consider the constant presence that these complex entities maintain on all aspects of global networked communication. Additionally, this system provides a very specific implementation of a more generalized system for creating audio based augmented realities. More possibilities exist for local, regional, continental, and worldwide installations using this system. I intend to further explore these variations in scale as well as new contexts that, with the help of this technique, can gain additional meaning.

References

Adenot, P. & Toy, R. (2016, November). *Web Audio API Editor's Draft 29*. Retrieved from https://webaudio.github.io/web-audio-api/.

Adenot, P. & Wilson, C. (2015, December). *Web Audio API W3C working draft 08*. Retrieved from https://www.w3.org/TR/webaudio/.

Carlile, S. (2011). Psychoacoustics. In Hermann, T., Hunt, A., Neuhoff, J. (Eds.), *The Sonification Handbook* (pp. 41-62). Berlin, Germany: Logos Verlag

Chion, M., Gorbman, C., & Murch, W. (1994). *Audio-vision: sound on screen*. New York: Columbia Univer-sity Press.

Fuller, M. (2005). *Media Ecologies*. Cambridge, MA: MIT Press.

Hansen, M. B. N. (2014). *Feed Forward: On the Future of Twenty-First Century Media*. Chicago, IL: University of Chicago Press.

Kahn, D. (2013) . *Earth Sound Earth Signal*. Oakland, CA: University of California Press.

Krause, B. L. (2011). *The great animal orchestra: finding the origins of music in the world's wild places*. New York: Little, Brown.

Neuhoff, J.G. (2011) . Perception, cognition and action in auditory displays. In Hermann, T., Hunt, A., Neuhoff, J. (Eds.), *The Sonification Handbook* (pp. 63-86). Berlin, Germany: Logos Verlag.

Serafin, S., Franinović, K., Hermann, T., Lemaitre, G., Rinott, M., Rocchesso, D. (2011). Sonic interaction de-sign. In Hermann, T., Hunt, A., Neuhoff, J. (Eds.), *The Sonification Handbook* (pp. 87-106). Berlin, Germany: Logos Verlag.

Thompson, M., & Biddle, I. D. (2013). *Sound, music, af-fect: theorizing sonic experience*. London: Bloomsbury Academic.

Thrift, N. (1997) . The still point: expressive embodiment and dance. In Pile, S. & Keith, M. (eds.), *Geographies of Resistance* (pp. 124-151). London, England: Routledge.

Walker, B. & Nees, M. (2011). Theory of Sonification. In Hermann, T., Hunt, A., Neuhoff, J. (Eds.), *The Sonifi-cation Handbook* (pp. 9-40). Berlin, Germany: Logos Verlag.

Author Biography

John Brumley is currently a PhD candidate in the department of Empowerment Informatics at the University of Tsukuba in Japan where his research focuses on using mixed reality systems to promote collaboration in physical space. Brumley received his MFA from UCLA in 2015 in the department of Design Media Arts, producing digital and physical work regarding contemporary virtuosity, localized collaboration, and digitally saturated adolescence. He holds a BA in music composition from UC Davis where he spent time as a radio DJ, improviser, and a practitioner of Sundanese gamelan music.