# Spatially Distributed Sound Computing and Rendering Using the Web Audio Platform

Lonce Wyse
Communications and New Media
National University of Singapore

lonce.wyse@nus.edu.sg

## ABSTRACT

Large multi-channel spatial audio systems have historically been play-grounds for universities and well-funded studios, but only a dream for independent composers. Similarly, "parallel computers" were locked in research facilities, where only a few musicians ever gained access to, for example, the compute power to convolve hundreds of separate audio streams with spatially-specific room impulse responses. Mobile devices in the hands of audiences can quickly configure themselves into such systems at very affordable (and distributed) cost and little effort, making powerful and expressive spatially distributed musical platforms accessible to anyone today. We describe some software systems and artistic works that have been developed recently to explore some of the spatial audio capabilities of the mobile device browser platform.

## Categories and Subject Descriptors

H.5.5 [**Information Systems**]: Information Interfaces and Presentation (HCI) – *sound and music computing*.

## General Terms

Performance.

## Keywords

Spatial Audio; Interactive Audio; Distributed computing; Audience participation; Web Audio API.

## 1. BACKGROUND

There is a long history of multi-channel sound systems in electroacoustic music. The Acousmonioium [3] was comprised of dozens of heterogeneous speaker spread over stage and auditorium controlled from a mixing console. “Diffusion” is recognized as an artistic practice in its own right, referring initially to performing at the mixing console to spatially render prerecorded music [4], and later to include possibly live electroacoustic music. The Phillips Pavilion at the 1958 World's Fair was fitted out with an estimated 350 speakers for the installation and performance of Edgard Vares’s Poème Élctronique [6]. The BEAST at the University of Birmingham was founded in 1982 by Jonty Harrison and is based on the Acousmonium model and is today typically configured as an 80 independent channel system. The Sonic Laboratory at the SARC is equipped with over 120 speakers that includes placement beneath the floor level [1]. Multi-channel speaker domes range from small to large, with the UCSB Allosphere being comprised of 140 speakers [2].

For all their high-fidelity and spatial resolution capabilities, these systems have several distinct characteristics that define the envelope of musical territory for which they were designed. First is the need to distribute high-bandwidth audio to possibly very many speakers at the same time – hundreds of channels of 48000 24-bit samples is a lot of data for a communication system. If each channel of data must be centrally computed under real-time control, then compute cycles becomes an issue. Traditional multi-channel speaker systems such as those mentioned above, as well as those found in movie theaters and homes, tend to be placed outside the perimeter of the audience, sometimes simulating movement through the audience with a variety of simple and complex processing techniques with various degrees of spatial accuracy. Finally, speakers generally have fixed spatial locations, and are literally tethered to their locations with wires. This makes them very difficult to reconfigure, not to mention physically move during performance (although this did not prevent Gordon Monohan from creating his (1982) *Speaker Swinging* piece).

Mobile devices such as phones and tablets can also be conceptualized and deployed as spatial audio diffusion systems. A concert hall for example, can easily and at no cost to a composer, be filled with many hundreds of sound sources that tend to be distributed in a conveniently uniform way throughout the audience. Despite the somewhat limited fidelity and volume potential, this diffusion system has advantages over the “heavy” systems described above for certain musical applications. The bandwidth and the CPU bottlenecks are considerably reduced if only control signals are communicated to the ad-hoc massively distributed computing nodes which then compute and render the audio. Logistics are simplified by the wireless communications allowing the speaker configuration to be customized for each performance venue and audience.

## 2. ADIFFUSION FOR DISTRIBUTED MUSICAL ENVIRONMENTS

ADiffusion is a platform for developing distributed sound works. It consists primarily of a server written in JavaScript using the node.js framework, and is modelled on a “chat room” architecture. The server is a very light-weight. It defines a small number of messages that it handles (e.g. the “subscribe” message sent when a client joins), but otherwise simply receives messages from a client and does nothing but distribute received messages to all the other


1st *Web Audio Conference*, January 26–28, 2015, Paris, France.

clients who have subscribed. The server also sends out periodic clock signals that clients can use for synchronization.

The clients use a communications library for creating and receiving messages consisting of a name and an arbitrary array of data wrapped as a JSON object. The clients also use the jsASound library [7] built on the Web Audio API for sound synthesis. The composer creates a web page for clients whose primary function is to respond to messages and create sound. The web pages can also make use of the screen and/or sensors for gathering data to create "controller" messages and for displaying information through channels other than audio (graphically, vibro-tactically).

The ADifussion system has been used in a variety of musical contexts as described next.

## 3. CONCERT VENUES

Mobile phones were used in a concert hall as a distributed speaker system as early as 2001 for Golan Levin's *Telesymphony* (2001). Levin and his collaborators had people install ring tones on their phones and performed the system by dialing numbers from the stage [5].

The more modern computational and network technologies simplify the implementation of this concept and enrich the musical possibilities. Levin had to install ring tones on audience phones prior to performance, and collect the phone numbers of the devices and associate them with seat locations. Thirteen years later, we can create arbitrary synthesis algorithms to run on the mobile devices, and define our musical communication protocol in whatever way suits our musical objectives.

ADiffussion was used in the concert hall during a GENUS Guitar Ensemble concert at the National University of Singapore in September, 2014. The piece *AME* by Kengo Momose starts off with rain-like sounds tapped with fingernails on the body of acoustic guitars. It climaxes with thunderous timpani accompaniment, and ends with music suggestive of the emergence of the sun following a storm.

Prior to the beginning of this particular performance of the piece, the audience was invited to turn on their mobile devices and to navigate their browsers to the concert website where they were requested to enter their seat number. A soft rain sound then began in the rear the auditorium. It slowly spread its way toward the stage over approximately 15 seconds when the performers began the piece with the guitar body tapping.

To visually accompany the thunder in the middle of the piece, the house lights dimmed and flashes of white light flickered from devices throughout the audience. When the piece reached its conclusion, sounds of a sunny day remained for a short time among the audience in the form of 25 different bird songs. Several audience members commented on the physically immersive sense of being in and among the distributed sound sources.

The way this piece was implemented in ADiffusion was with two types of clients, each with their own web page - one for audience members, and one for a special client called the "conductor." All joined the message-exchanging chatroom on the server, but the conductor was the only source of messages during the piece. The conductor had a simple web-based graphical interface sending messages for the 3 separate sections of the piece where audience devices were enlisted for sound or light.

The strategy for clients was also straight-forward. Although it would have been possible to independently address each device by seat number since this data was provided from clients to the server, this level of complexity was unnecessary. Audience devices all received the same messages, but were programmed to respond differently by adjusting their sound synthesis algorithms according to their seat number. Thus, the rain moved across the audience in response to messages indicating location, and the density of the randomly distributed bird sounds was controlled based on a message containing the number of participants in the audience.

## 4. GALLERY VENUES

A gallery installation creates an entirely different set of challenges and opportunities for distributed sound pieces on mobile devices. In this environment, people are constantly on the move as they explore the gallery space, so position-specific sound design is (currently, at least) far more challenging. Furthermore, in a concert hall, composition and diffusion strategies can depend on a reasonable estimate of the number of participants. Even if only a portion of the audience actually engages, the participating devices are statistically likely to be roughly evenly distributed throughout the audience. However, a gallery installation work needs to work with audiences ranging from 1 to 20, or possibly many more. It is the smaller numbers which are most difficult for pieces designed to be distributed. If only a portion of the gallery visitors engage with their mobile device, for example, the numbers may be too small to assume anything about their distribution through the space.

In November 2014, *States of Diffusion for n+1 devices* was installed in a gallery for the ACM Multimedia exhibition (Wyse, 2014). The work is built with two different types of client browser sound sources – one for the gallery which renders sound on a pair of high-fidelity stereo speakers, and the other for visitors to the gallery with their mobile devices (Figure 1). A "conductor" client is used in this work, too, to step the piece through a randomized sequence of one-minute sonically distinct sections. The "+1" in the title represents the "gallery" device. The presence of the gallery sound system means that even for a solo gallery visitor, there will always be a sonic partner to make the collaborative and distributed aspects of the piece work.

*States of Diffusion* takes advantage of the naturally moving sound sources for its sonic effect. For example, one of the six sections consists of a timbrally rich drone which is very slightly mistuned among the various devices. Slight mistuning would normally cause an audio "beating" effect, but since the sources are distributed and moving, the beating is also spatial and changing according to the visitor's gallery behavior. *States of Diffusion* also takes advantage of device sensors for compass direction and rotation, but only for interacting with local synthesis computation.

## 5. CHALLENGES FOR DISTRUTED SOUND ON THE WEB PLATFORM

Some obstacles exist for designing audio experiences such as those described above just because of simple, hopefully temporary, issues with the current state of the mobile browser platform. For example, there is currently no way for browser program developers to keep mobile devices from going to sleep (though users can do so through device settings). Once devices sleep, they don't accept network messages (to turn sounds off, for example). Another issue in this category is the inability of programmers to prevent incoming message and phone call notification. This can be particularly disruptive in a concert hall environment. Once this issue is addressed, it would be possible to engage audience-held devices through browsers in a wide variety of performance environments such as theatre and film. Browser programmers are also unable to consistently prevent screen rotation. This is a problem whenever device motion is an integral part of the user experience (e.g. in gameplay or musical instrument interfaces). These three things (sleeping, notifications, and screen rotation) could easily be permission-granted on a per-webpage basis and would make the browser platform much more usable for collaborative audience experiences.

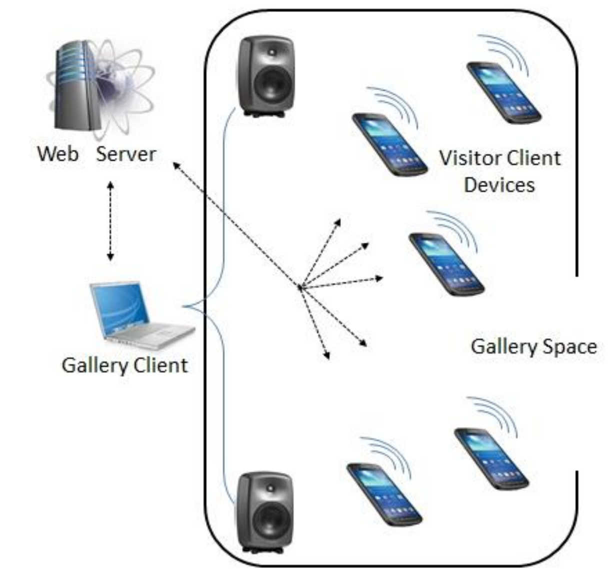


**Figure 1. Clients point browsers to the web server to access the *States of Diffusion* website. The Gallery Client is always present so even solo visitors have a collaborative and spatial sound experience.**

Timing and synchronization is a deeper problem. There are many kinds of timing issues, such as message transport latency and jitter. Just to take one timing issues as an example, if clocks on various devices are not synchronized to within a few tens of milliseconds, it is not possible to render a typical drum set "rhythm" with instruments distributed across different devices. It is possible to create perfectly periodic distributed event patterns, but not to specify exactly the timing of the desired events within the repeating pattern since this scenario depends on the timing relationship between different devices. Several strategies have been proposed for synchronizing clocks on mobile devices, but there is no standard way to do this yet. Solving this issue will open up vast new musical possibilities for the distributed browser platform.

Finally, testing during development for this platform is challenging. While we extolled the virtues of the spontaneous formation of the distributed system at performance time above, the performance platform unfortunately doesn't exist at development time.

It is possible to simulate an audience by automating the opening of a number of browser windows on a single computer, each directed to the appropriate URL for their role in a piece. However, a single PC does not have nearly the compute power of a hundred mobile devices, so testing is limited to a restricted number of devices. Secondly, running all the clients on a single computer does not present the spatial characteristics of the performance platform. Upon further reflection however, perhaps this kind of limitation is not so different from composing an orchestral work using only a piano.

## 6. SUMMARY

New musical possibilities, particularly for spatial audio, are enabled by the ubiquity of mobile devices, their computational power, the ease of audience participation through browsers, and emerging browser specifications such as the Web Audio API. Some of the possibilities were explored with examples of recent pieces built on the ADiffusion communication platform, and the ASound modelling and synthesis library.

## 7. ACKNOWLEDGMENTS


The software used to construct the works described in this paper are open source and available on GitHub and at http://anclab.org/Software. Thanks to contributors to the open source Web Audio ecosystem upon whose shoulders the jsASound library and *States of Diffusion* are built.


## 8. REFERENCES


[1] Alcorn, M. and Corrigan, C. 2003. Sonic Arts Research Centre (SARC). *Proceedings of the... International Computer Music Conference* (2003).

[2] Amatriain, X. et al. Experiencing audio and music in a fully immersive environment. *Computer Music Modeling and Retrieval. Sense of Sounds*. Springer. 380–400.

[3] Bayle, François 2007. Space, and more. *Organised Sound*. 12, 03 (2007), 241–249.

[4] Harrison, J. 1998. Sound, space, sculpture: some thoughts on the "what", "how" and "why" of sound diffusion. *Organised Sound*. 3, 02 (1998), 117–127.

[5] Paul, C. 2003. *Digital Art*. Thames & Hudson.

[6] Trieb, M. 1996. *Space Calculated in Seconds: The Philips Pavilion, Le Corbusier, Edgard Varèse*. Princeton University Press.

[7] Wyse, L. 2014. Interactive Audio Web Development Workflow. *Proceedings of the 22nd ACM International Conference on Multimedia* (Orlando, Florida, Nov. 2014).