Merging Places: A Real-Time Distributed Live Reverberation Chamber

Austin Franklin
Mälardalen University
Västerås, Sweden
austin.franklin@mdu.se
The Royal College of Music
Stockholm, Sweden
austin.franklin@kmh.se

Daniel Hedin
Mälardalen University
Västerås, Sweden
daniel.hedin@mdu.se
Chalmers Univeristy of Technology
Gothenburg, Sweden
daniel.hedin@chalmers.se

Rikard Lindell
Mälardalen University
Västerås, Sweden
rikard.lindell@mdu.se
Dalarna Univerity
Falun, Sweden
rli@du.se

Henrik Frisk
The Royal College of Music
Stockholm, Sweden
henrik.frisk@kmh.se

***Abstract*—We present Auxtrument, a prototype instrument that allows audiences to experience the acoustic qualities of remote locations. Using the metaphor of a reverberation chamber, artists send signals from the mixers, bus, or aux via the Auxtrument's network connections from a concert hall to a few different locations. At each location the signal is played through loudspeakers and captured, colored by the acoustics and noises of the place, via a stereo or ambisonics microphone. The signal is sent back and played for the audience in a surround sound system conveying the spatial qualities of the places. The Auxtrument allows us to merge and layer the different locations in the concert hall. However, this arrangement places great demands on the network. The audio signals need to be high-resolution to preserve the inherent quality of the sounds, which creates large streams. The system must be equipped to work over different types of networks, and to enable any location, the system must work with mobile devices. After ruling out several commercial and open-source solutions, we built the Auxtrument with web technologies: mainly node.js, WebSockets, WebRTC, and WebAudio.**

***Index Terms*—merged reality, IoT, distributed performance**

I. Introduction

Location, space, and time are just a few elements of a musical listening experience. However, they are as inseparable from the emotional and physical impact on the listener as the type of reed or string the performer uses to conjure a note. The significance of these elements has been acknowledged since ancient times, with civilizations leveraging natural acoustics to enhance auditory experiences. Paleolithic communities in France harnessed the reverberations of caves like Lascaux [7]. In Western architectural traditions, ancient Greek amphitheaters such as those in Miletus, Rhodes, Siracusa, and Epidaurus were designed to optimize the projection of performers' sounds, showcasing an early understanding of spatial dynamics in music [13].

Music is inextricably bound to the wider auditory world, since it sounds within it. With an ecological take on music perception, Eric Clarke states that music "incorporates environmental sounds into its own material, and... takes on fluid relationships with the physical and social spaces that it occupies" [4]. When discussing acoustic spaces, it is possible to consider the space as an instrument itself, a resonant chamber, that co-evolves with the music. Today, using a site-specific paradigm of composing is not uncommon. These pieces are written to resonate harmoniously within a particular locale's unique physical properties [2].

The term 'acoustic space,' given by Marshall McLuhan, refers to "a sphere whose focus or center is simultaneously everywhere and whose margin is nowhere... Acoustic space is dynamic; it has no fixed boundaries. It is space created by the method or process itself" [8]. The modern soundscape is an adaptation of McLuhen's acoustic space, established at Simon Fraser University in 1970 by the World Soundscape Project. The Project's interest in the experiential qualities of sound relating to location has inspired audio walks and soundscape field recordings, both of which attempt to harness the power of location, space, and time to influence sensory experience and perception [10].

Like McLuhen, Pettitt (2007) predicted that digital transformation would change the view of the work. The COVID-19 pandemic exacerbated the transformation where concerts and musical meetings moved online, but the experience of the performance was often deprived of qualities that constitute liveness [3]. For most of these performances, the view of an acoustic space (traditionally a concert hall) was reduced to a mere stereo or even mono audio signal.

Recent advancements in information and communications technology (ICTs) have enabled people who stay or move in space to interact with each other or with the places where they are immersed. They can directly transmit information related to the physical space or socially share information received from sensors located not only in the main nodes but also along and around in-between spaces. With audio specifically, these advancements have the potential to break down barriers to access and representation within live sound production. This relationship between people and places within an environment redefines the mainstream internet of things (IoT) concept as that of an internet of places concept [9].

We have tried several solutions open-source and commercially available solutions, such as SonoBus and JackTrip. However, these were found to be inadequate for our purposes in key ways. JackTrip's desktop app and CLI do not work

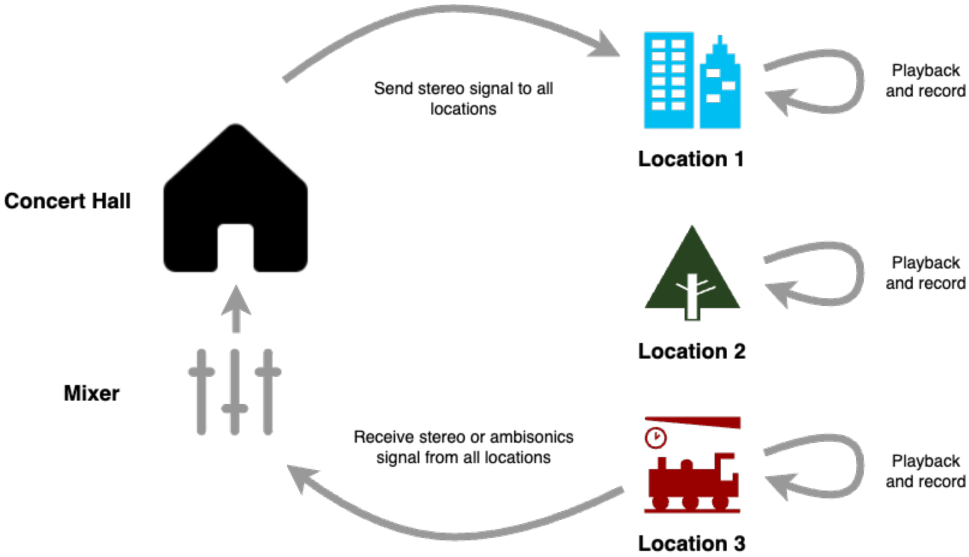

Fig. 1. The Auxtrument Instrument Design

with mobile devices. SonoBus showed initial promising results – it allowed sending and receiving multi-channel signals and supported mobile devices. However, there were issues with connecting over different networks and between mobile phone carriers. Although SonoBus uses peer-to-peer connections, there is no interface for selecting the port used between participants, so troubleshooting and establishing a connection is difficult if you are behind a Network Address Translation (NAT) service and port forwarding is unavailable. For these reasons, we have decided to build the Auxtrument as a web application using node.js, WebSockets, WebRTC, and WebAudio, among other libraries that facilitate the use of these tools.

II. Instrument Design

The Auxtrument is like a distributed reverberation chamber that merges acoustic qualities of remote places. A concert hall performance is captured in real-time and sent to a number of satellite locations. The signal at each satellite location is played through loudspeakers and re-recorded, colored by the physical properties of the acoustic space, via a stereo or ambisonics microphone. The captured signals are transmitted back to the concert hall and diffused in a surround sound system, which conveys the unique spatial qualities of the space. An Auxtrument with three (3) satellite locations is shown in Figure 1.

We identified three (3) key demands that must be satisfied prior to development for high-fidelity user experience: 1) the audio signals need to be high-resolution to preserve the inherent quality of the transmitted sounds and accurately relay spatial properties of the acoustic spaces; 2) it needs to be easy to set up and use, and work over different types of networks; and 3) to enable any location, the system must work with mobile devices.

III. Network Configuration

The Auxtrument uses UDP over a peer-to-peer connection that requires at least one (1) of the participants to be reachable from the other. In our case, with satellites being mobile devices and the concert hall being behind a NAT, neither is able to initiate the connection. The same type of problems also manifest in the presence of firewalls that prevent the external network from initiating communication with hosts on the internal network.

Several solutions for initiating connections between peers exist due to the prevalence of such network setups and peer-to-peer based applications. Generally, most rely on a technique known as hole punching. Somewhat simplified, hole punching is based on indirectly manipulating the state of the firewall or NAT by initiating communication from the inside of a network in such a way that the initiator is exposed on the external network [6].

For the Auxtrument we have settled on using WebRTC.[1] WebRTC is a standard for real-time communication on the web with a specific focus on audio and video. It relies on STUN (Session Traversal Utilities for NAT) servers and ICE (Interactive Connectivity Establishment) to find ways for two (2) peers to connect directly to each other (Figure 2). ICE uses the STUN server to probe various ways to connect to the clients from the outside – to punch holes. The results are transmitted back to the peers to allow them to establish a peer-to-peer connection. However, if the network is explicitly configured to prevent peer-to-peer connections, WebRTC uses TURN (Traversal Using Relays around NAT) servers that externally relay communication between the peers. In the setup of WebRTC, peer identifiers (that allow peers to connect to each other) must also be distributed between the peers – a process known as signaling [11].

[1] webrtc.org

From an implementation perspective, the Auxtrument is a WebRTC based web application using express.js[2] running on top of node.js.[3] The peer.js[4] library is used to help simplify the process of setting up a WebRTC connection. This stack helps streamline and simplify the number of actions a user must perform to establish a connection. Users simply navigate to the URL and select whether they are the concert hall or location peers. The signaling and peer-to-peer initiation is automatically handled in the background, and the audio input and output devices default to those of the systems running at each place.

IV. Web Audio

The Auxtrument uses the open-source and royalty free Opus codec. The Valin and Brans' WebRTC Audio Codec and Processing Requirements[5] mandates that all WebRTC-compatible browsers must support Opus. Its scalability and flexibility are useful when dealing with audio streaming that has varying degrees of complexity, and it is robust enough to deliver high quality transmissions. Opus supports constant and variable bitrate encoding from 6 kbit/s to 510 kbit/s (or up to 256 kbit/s per channel for multi-channel tracks), frame sizes from 2.5 ms to 60 ms, and five (5) sampling rates from 8 kHz (with 4 kHz bandwidth) to 48 kHz (with 20 kHz bandwidth, the human hearing range). The Auxtrument uses Opus with a sampling rate of 48 kHz to match that of most web browsers to prevent resampling at any stage in the transmission process, along with variable bitrate encoding.[6]

The recording for both the satellites and the concert hall is handled using the web browsers' `getUserMedia()` javascript function of the MediaDevices API, which returns an audio stream from the device's selected audio input device. At the same time, audio playback uses the Web Audio API.[7] Once the audio has been received from all locations, the concert hall implementation parses the channels from the streams of each location and merges them into a single multi-channel stream. For instance, three (3) locations with three (3) stereo streams are merged into a single stream with six (6) channels that is sent to a multi-channel audio device and controlled via a mixer.

When setting up the aforementioned WebRTC connection, the protocol requires information about the stream configuration. Session Description Protocol (SDP) munging is required to configure a media stream to contain more than the default single channel. The SDP is the standard that describes media communication sessions, and it contains the codec parameters, source address, and timing information of audio and video streams [12]. Capturing and streaming a stereo signal is straightforward and requires only forcing `stereo=1` to the description.

[2]expressjs.com
[3]nodejs.org
[4]peerjs.com
[5]https://datatracker.ietf.org/doc/rfc7874/
[6]https://datatracker.ietf.org/doc/html/rfc6716
[7]https://www.w3.org/TR/webaudio/

V. Demonstration

We propose a demonstration of the Auxtrument in concert, where music performed in a concert hall is transmitted over a network, filtered through acoustic spaces, and sent back to the concert hall. Like reverberant cathedrals or caves, we take advantage of the natural acoustics of far-away spaces and bring them to an in-person audience. Through layering and merging, we are able to create new hybrid spaces with artificial properties that reshape sensory perception.

The Auxtrument demonstration will consist of a concert of fixed media works created by the authors that are streamed to locations in Sweden near the conference site and cities such as Stockholm, Uppsala, and Västerås. Onsite conference participants will be able to experience the natural acoustics and environmental artifacts of these spaces via headphones. The Auxtrument equipment requirements (to be provided by the authors) consists of a laptop running the web app, an audio card, and headphones. The offsite equipment at each location consists of a mobile phone, portable speaker, stereo pair of microphones, and an iOS and Android compatible mobile audio card.

Drawing upon Erin Clarke's framework of musical space, we can conceptualize the audience's perception of sound within the concert hall as being deeply intertwined with their bodily presence and spatial orientation within the environment, where perception is defined as, "the awareness of, and continuous adaptation to, the environment" [5]. As we are interested in user perception, we will evaluate Auxtrument by conducting participant interviews during the demo. This information will help us to better understand the complex interplay between auditory perception, spatial cognition, and emotional resonance.

VI. Conclusion

The Auxtrument is a prototype instrument that allows artists to send audio from a concert hall over a network to several remote locations. These locations playback and record the sound, colored by the space, with a stereo or ambisonics microphone. The signal is transmitted back to the concert hall and diffused in surround sound for a live audience, conveying the spatial properties and ambient noises of each location. The Auxtrument is built primarily using web based technologies such as node.js, WebSockets, and WebAudio; it has an easy to use interface and works with different types of networks; and it works with mobile devices. The experiential qualities of this instrument were discussed along with the technical challenges and features of the design.

Acknowledgment

Information Retrieval in Embedded Systems for Audiovisual Artistic Processes (IRESAP) is supported by The Knowledge Foundation (KKS), with additional support from Ericsson, Spotify, and teenage enginnering.

References

[1] Adhikari, M., Hazra, A. (2022). IEEE Communications Standards Magazine, 6(1), 67-74.

[2] Bagenal, H. (1927). Influence of buildings on musical tone. Music Letters, 8(4), 437-447.

[3] Benford, S., Mansfield, P., Spence, J. (2021, May). Producing liveness: The trials of moving folk clubs online during the global pandemic. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-16).

[4] Born, G. (Ed.). (2013). Music, sound and space: Transformations of public and private experience. Cambridge University Press.

[5] Clarke, E. (2005). Ways of listening: An ecological approach to the perception of musical meaning. Oxford University Press.

[6] Ford, B., Srisuresh, P., Kegel, D. (2005, April). Peer-to-peer communication across network address translators. In USENIX Annual Technical Conference, General Track (pp. 179-192).

[7] Kolar, M. A., Fritz, C., Tosello, G. (2022). Acoustics in Music Archaeology: Re-Sounding the Marsoulas Conch and Its Cave. Acoustics Today, 18(2), 52-61.

[8] McLuhan, M.; Powers, B.R. The Global Village: Transformations in World Life and Media in the 21st Century; Oxford University Press: New York, NY, USA, 1989.

[9] Morandi, C., Rolando, A., Di Vita, S. (2016). From smart city to smart region: Digital services for an Internet of Places (p. 103). Cham: Springer.

[10] Saunders, A., Moles, K. (2016). Following or forging a way through the world: Audio walks and the making of place. Emotion, Space and Society, 20, 68-74.

[11] Rodríguez Baquero, D. (2021). Analysis of WebRTC signaling.

[12] Roy, R. R. (2018). Handbook of SDP for Multimedia Session Negotiations: SIP and WebRTC IP Telephony. CRC Press.

[13] Thün, G., Velikov, K., Ripley, C., Sauvé, L., McGee, W. (2012). Soundspheres: resonant chamber. In ACM SIGGRAPH 2012 Art Gallery (pp. 348-357).