Solving for Latency in Distant Musical Performance

The Velocity of Sound

Written by Warick Pond
Executive Director,
ASU Enterprise Technology

“In the new era, thought itself will be transmitted by radio.” - Guglielmo Marconi

While COVID-19 forced many employees to work from home, many restaurants, campuses and music venues closed as everyone scrambled to pivot to virtual or reduced contact services. During that period, many of us had the benefit of being able to adjust our working and learning environment from in-person to virtual. In fact, a survey by Owl Labs indicated that during COVID-19 close to 70 percent of full-time workers were working from home, where they continue to interact with cohorts, clients, and customers from ad hoc offices created during the pandemic time. These “off-site” employees are able to collaborate utilizing video conferencing, file sharing, messaging app technologies and more.

One area of learning and work that has not been able to adjust to virtual participation and collaboration has been collaborative popular music creation and performance. Sure, we’ve all experienced virtual musical production videos with multiple performers seeming to play together all at once. However, those were all created using individual performances that were later synchronized and placed in window boxes in post-production, to give the illusion of a live group performance.

With 2.12 million people in the music workforce, which has a growth of 3 percent annually, this is a significantly disenfranchised workforce that does not have equitable access to collaborate virtually in real time. ASU Enterprise Technology sought a proof-of-concept to create a near real-time collaborative virtual performance.

The Problem

At the core of the issue are the physics of distance and time. In any network the time delay (latency) between participants gets worse as the distance between participants increases. If the goal is to connect people everywhere to instantly perform together, deepen collaboration, develop music and create engagement from viewers, then a real time experience is required. With highly syncopated music such as pop, hip hop, blues and rock, where matching beats and tempos are required, a delay of 10 milliseconds (ms) or less usually will not affect the performance. Above 10-12ms, the latency starts to become a noticeable distraction with a degradation of the experience. Musicians performing via streaming online must invest thousands of dollars into instruments, hardware, software and sound-creation devices, and then educate themselves (beyond the mastery of their instruments) in the technical concepts required to reduce audio and visual latency. Converting analog instruments to digital requires a high degree of knowledge and understanding of the ramifications and intricacies of the inherent latency of digital effects processors, reducing buffer size in their digital audio workstation (DAW) and monitoring CPU usage, latency in headphones and real-time audio suite (RTAS) audio plug-ins. Solving the problem of latency includes aligning to ASU's Charter of inclusion, meaning identifying a solution that is accessible with a strong ease of use.

Building the Band

The challenge required the right mix of experts and technology partners, including an ASU technologist passionate about the problem, network architects, video and audio-conferencing experts and a music industry expert in musical performance hardware and other equipment. Enterprise Technology utilized our existing partnership with Verizon and their 5G infrastructure as well as ASU's Zoom Innovation Lab. We also partnered with a former senior Vice President at Fender Guitars. An additional hardware provider was included through Elk Audio.

The gathering of the "band" members resulted the following make up:

Partner	Band Member	Role
Arizona State University, Enterprise Technology	Warick Pond, Executive Director	Executive Director and musician
Arizona State University, Enterprise Technology	Olivia Hernández, Creative Manager, Learning Futures	Project Coordinator, stage manager
Verizon	Jim Pfleger, Associate Fellow	Technical lead, network expert
Music Industry Professional	Ritchie Fliegler, Former Sr. VP at Fender	Musician and hardware expert
Elk Audio	Michele Benincaso	Hardware and OS provider

Packing the Van with Equipment and Devices

There were several different and important devices used for these tests. The project ambition of ensuring ease of use and accessibility meant that these devices needed to be affordable, easily deployed, and accessible. As musicians, we can geek out about all the different tools and equipment we use and why. This report will focus on the unique assemblage used to perform in a virtualized environment and not the specific cords, guitars, mics and pedals used.

The unique hardware we utilized included the Inseego MiFi® X Pro 5G mobile hotspot router, Verizon Fixed Wireless Access, Elk Audio Bridge and the Elk Audio LIVE Studio.

The Inseego MiFi® is unique in that it has an ethernet tethering port. The hotpot router is built with a 5G sub-6/mmWave or sub-6 network technology. It is conveniently small at 5.9” x 2.9” x .74”. Verizon 5G Business Internet is a plug-and-play, fast and scalable wireless connectivity solution. It includes an all-in-one router that provides both wifi and ethernet ports for connecting a home or business location to Verizon’s high-speed, low-latency 5G Ultra Wideband network.

Elk Audio Bridge includes 2 unbalanced line outputs as well as ADAT, S/PDIF and MIDI. Analog inputs include combo connectors for XLR and TRS. Outputs have both 1/4 inch and 3.5mm. USB-C charging and an ethernet tethering port are included. The Elk Audio Bridge must be on the same internet connection that the Elk Audio LIVE Studio is connected to via the ethernet tethering port to the router. Elk Audio LIVE Studio is a browser-based session control room. It enables control of levels and sends two channels of audio (instrument and vocals) and receive the same.

Unique Equipment:

- **Inseego MiFi® X Pro 5G mobile hotspot router**
- **Verizon Fixed Wireless Access**
- **Elk Audio Bridge**
- **Elk Audio LIVE Studio**
- **Zoom Video Communications**

Jamming Sound Waves with Millimeter Wave

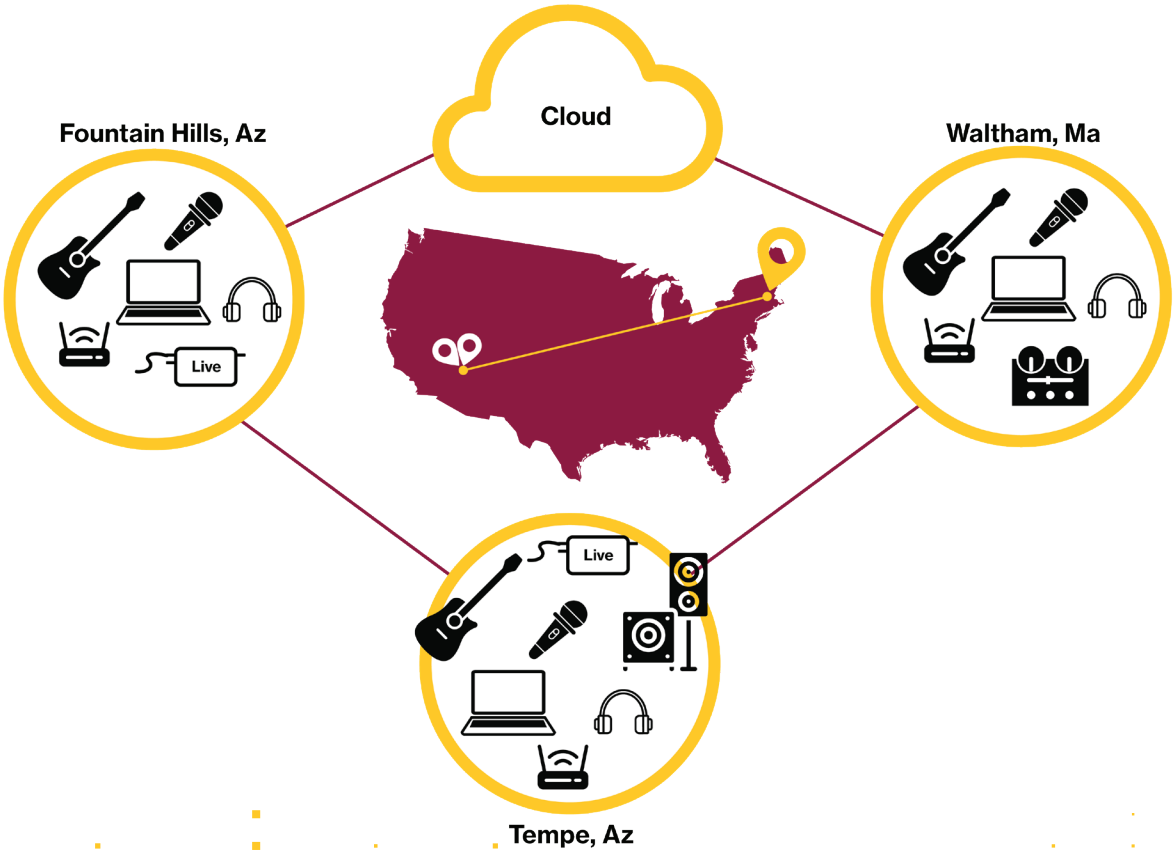
Enterprise Technology is excited to say that we successfully held a synchronized virtual jam session at a distance of 2,632 miles away between participants. Our setup included two performers in Arizona and one in Waltham, Massachusetts.

The utilization of the 5G technology made the paramount difference in performance and ability to synchronize in real-time enough to collaboratively play a 90 beats per minute (BPM) blues riff and solo.

The test was part of ASU's Realm 4: Education Through Exploration Summit, where both Warick Pond and Ritchie Fliegler anticipated demonstrating the difficulties of virtual popular music performance play.

The results startled the team, as a week prior the two performers had tested at 421 miles with heavy latency. Those tests were done without the 5G technology. The results of 2,632 miles away indicate a proof of concept for virtualized performance and collaboration for musicians in real-time.

The applications are numerous. At present, the popular music industry requires in-person collaborations or performances. Imagine the musician who no longer must relocate to musical city hubs across the world just to work or study with other musicians. Connecting musicians from across the world increases creativity and productivity and reduces the human footprint of musical production.



Hitting the Whammy Bar

While latency and distance were the most crucial issues for audio when doing live broadcasts, especially when the peers all need to be synchronized to a rhythm or pulse, it didn't stop there. A number of other impediments were met along the way in proving out a low latency real-time musical collaboration.

Single Threaded Use Case: The tools typically used for virtual collaborations prioritize features differently based on the use case anticipated by their designers. For instance, Zoom's use case of ensuring a participant can be heard includes the reduction of background noise, or ensuring that two participants are not able to talk over each other. This practical feature for conversation makes it difficult or impossible for musicians to play together at the same time.

Ease of Use: On the team are experienced experts in technology and musical performance. However, even with that high degree of practical knowledge, the team still found some tools cumbersome and difficult to use, having to fiddle with settings, hardware inputs and connections and configuring network preferences. An experience that requires that level of technical knowledge poses a barrier to accessibility at scale.

Inability to Stream: The tools we used for the proof of concept did not enable the ability to go live on the internet. The inability to simultaneously broadcast in real time reduces the use cases for many musicians as to why they would ever utilize virtual capabilities.

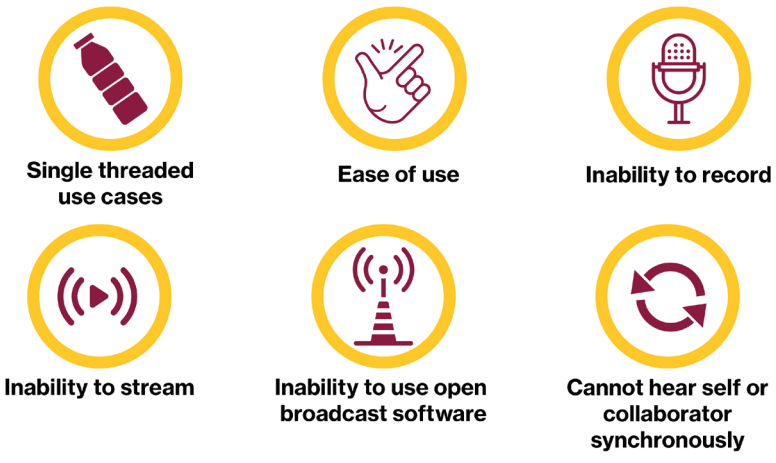
Inability to Record: Some tools enabled recording, others did not. Again, this simply reduces the use cases for musicians and collaborators. An additional tool would be required.

Inability to Use OBS: Open Broadcast Software would enable the ability to use an open source software for video recording and live streaming to online providers.

Synchronous Hearing: Participants are unable to either hear themselves or their musical collaborator. A bias toward using headphones exists with many tools.

Videoconferencing tools assume cross talking – unwanted signals in a communication – exist and have features to remove cross talking due to the use cases they are built to provide. Much of the collaboration between musicians includes conversation and cross audio sharing, being able to hear a partner's contribution while also hearing one's own musical contribution at the same time or different times. This is crucial especially in educational, improvisational or "free-form" settings. The assumption of cross talk existing prevents cross audio sharing. This also includes watching visual cues inclusive of the point of a hand, lift of an eyelid, tempo-synchronous body movements and more. Ensuring audio is synced with video is critical.

Impediments

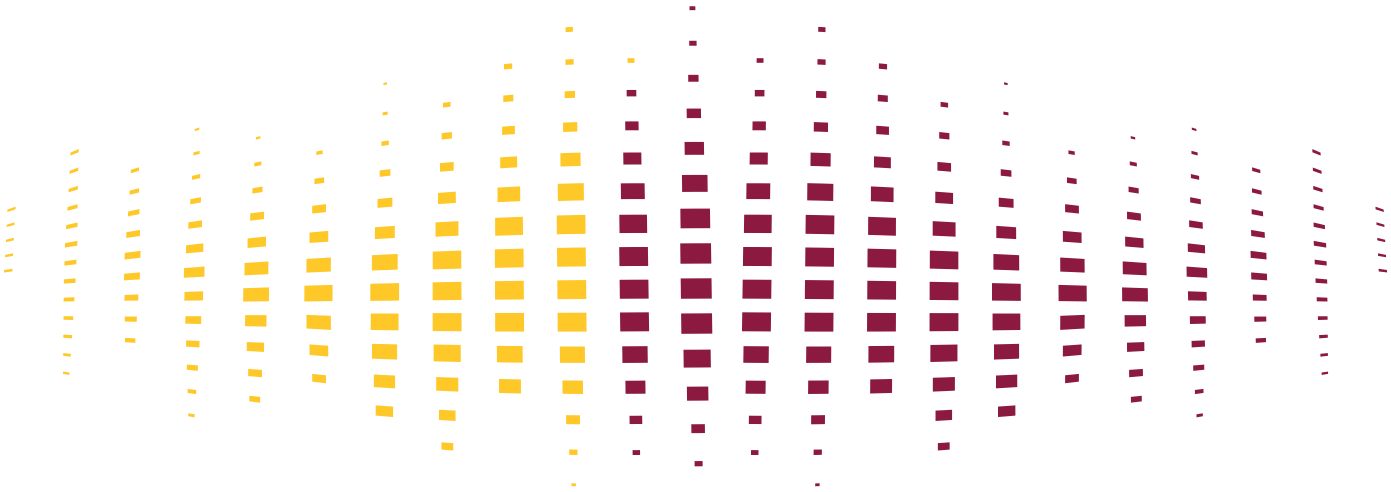


Next Gig

While these early results are promising the team would like to continue proofs of concept utilizing dedicated servers, virtualized avatars that create other world performances and unique fan-based experiences that all drive to new opportunities in the performance arts.

“As the realities of musical performance continue to be redefined, moving from theatres, stadiums, clubs and coffee houses to digital platforms, it points out the glaring inability of these virtual stages to allow for real-time, interactive group collaboration,” Fliegler said. “Solving this issue will literally change the way groups of musical and other rhythm-based performers coalesce and present their art.”

The team is seeking collaborators and sponsors for the work. For further details, please contact Warick Pond at warick.pond@asu.edu.



ASU Enterprise Technology
Arizona State University

verizon

Elk LIVE