# Telejam: From Low Latency to No Latency

Mark Goldstein
San Francisco, CA
kvetch@att.net

Dr. Michael McNabb
San Francisco, CA
michael@mcnabb.com

Boris Smus
Seattle, WA
boris@smus.com

## ABSTRACT

In this paper we describe Telejam, a novel web app for networked musical performance that connects players worldwide and allows them to play together in synchrony without regard to the time delays due to the distances between them.

## 1. INTRODUCTION

Sending audio across long distances takes time. Popular music jamming programs attempt to minimize the delay between players, but ultimately they fail at sufficiently large distances. In addition the two commonly-used architectures, many-to-many peer links, or server-based hub and spoke, cannot ensure that the delays between any two players are the same for everyone, so each player hears a slightly different mix. This creates more problems as the number of players increases.

Telejam was developed to let musicians play together with each other no matter how large the distance between them. Its design disregards the audio delay between musicians who can be very far away from each other.

Telejam was also designed to be easy to use. It is a web app that runs on the Chrome or Edge browsers, it requires no special computer skills, no special hardware, no software installation, and no user parameter setup. A wired ethernet connection is not necessary,  it can use wifi or a mobile hotspot, wired or wireless headphones and mics. The user interface is as simple as possible. The players' technical participation is deliberately minimal; they concentrate on playing and listening to each other.

## 2. SYNCHRONIZATION WITH TELEJAM

Telejam uses WebRTC to stream audio between players, and the Web Audio API to route audio within each player's machine.

Telejam is based on a peer-to-peer daisy chain. Each player connects to at most two other players: the one preceding them and the one following.

The chain can run in two jamming modes. In TALK mode  the audio connections between adjacent players is bi-directional. In PLAY mode audio flows only in one direction, from the first to the last player. A player receives an input stream from the preceding player, and sends an output stream to the next player. In STOP mode audio does not flow between players at all.

The unidirectional PLAY mode creates a zero-latency synchronized mix by performing a series of overdubs. Each player hears the incoming audio stream and plays along. Their input is mixed with the incoming stream, which guarantees synchronization, and the mixed stream is sent to the following player. In this way, a full performance is assembled, one player at a time. At the end of the chain the final mix can be live streamed to an audience and recorded.

Using a one-way stream eliminates the perception of latency. The players are not aware of the audio delay between them. The one-way stream involves a tradeoff: In exchange for precise synchronization each player only hears the mix from the peers on their left. The first player performs solo and can act as a de facto timekeeper or concertmaster. Successive players hear more and more parts as the mix accumulates and take their lead from what comes before. Everyone reacts in the moment to what they are hearing.

In TALK mode audio flows in both directions and is mixed with the player's input before being passed on. So players hear all of their peers on the left and right. This provides an unsynchronized, all-way conversational ability.

Since synchronization in PLAY mode is very accurate, a delay in the local system's audio path can be noticeable. The time between hearing the incoming stream and receiving the local player's simultaneous input should be as small as possible. Our goal is to detect this delay and account for it, so Telejam can accept all machine configurations including sound cards with large buffers, slower audio interfaces,  and other slow peripheral devices such as USB mics and Bluetooth headsets.

To measure the system delay we record a take in PLAY mode. The first player plays a steady pulse and each player claps along with it, one player at a time. The waveform of the recorded mix reveals any offset between the incoming stream and the player's input which was performed and mixed simultaneously.

Our mixing algorithm inserts a delay line between the player's incoming audio stream and their local input. The delay can be set to account for the system delay. This is a standard technique used in many digital audio workstations.

The calibration only needs to be done once. The system delay is solely a function of the local machine configuration and remains the same as long as the configuration is not modified. We believe calibration can be automated and this is a work in progress.

## 3. FOLLOW THE LEADER

Beyond synchronizing distant players, Telejam has additional features to make a musically productive and enjoyable experience. Telejam assigns special functions to a leader, so the other players are spared any technical responsibilities and can concentrate on listening and playing.

*Web Audio Conference WAC-2022, July 6–8, 2022, Cannes, France.*

Every Telejam session must have at least one player who is a designated leader. A leader can occupy any position in the chain but will usually be the first player. The leader arranges the order of the players and toggles the chain between STOP, PLAY, and TALK modes.

A leader can also set each player's input level remotely. In some cases, a second leader can occupy the last position in the chain in order to monitor the mix and set levels on the fly. Another possibility is to set levels using the loopback feature.

Although the first player in the chain has no incoming audio stream, the mix coming from the last player while in PLAY mode is sent back to the first player, who can monitor it. Of course this loopback signal will be delayed by the total transmission time of the chain. A slider controls the loopback volume, so the first player can hear the mix or mute it. The leader in the first position can listen to the mix and set the input levels to balance the ensemble before starting a performance, and then mute the loopback to start and participate in the performance.

## 4. RECORDING

Telejam can make recordings while in PLAY mode. The leader starts and stops recording takes and plays them back to the group while in TALK mode, so everyone can listen and comment. This provides immediate feedback and the opportunity to play and iterate, as in a recording session.

Telejam can also record the individual tracks played by each musician. Telejam uploads all recordings (mix and tracks) to a cloud database. The tracks can be downloaded later and remixed to create an alternate performance.

Since WebRTC transmits compressed audio, the final performance and its recording are compressed. However, the individual player tracks are recorded in-place at each player's machine and are uncompressed wav files. This offers the ability to remix the performance with  lossless audio.

## 5. AUX CHANNEL

In our first implementation, the peer-to-peer WebRTC connection was a mono stream, it only contained the ever-growing mix. By upgrading to a stereo stream we can add a second auxiliary channel for special use cases. We accumulate the performance mix in the left channel. The leader controls the input to the right "aux" channel which is passed down the chain unchanged.  The other players hear a mono mix of both channels. At the end of the chain, the full stereo signal can be recorded or broadcast, or the aux channel can be removed.

The aux channel can be used in different ways. It can provide a simple metronomic click or a pre-recorded backing track that holds the ensemble together so players can stop and restart at will without risking breaking synchronization for players further down the chain. Using the aux as a backing track, a choir can sing along to a pianist and the accompaniment can be removed at the end to produce an a cappella performance. The aux channel can also be used to layer performances recursively. A group can record a take, then use the just-recorded mix as the aux input for another take.

## 6. TELEJAMMIN'

We started using Telejam in June 2020 and have been experimenting with it weekly since then. Over time we simplified the players' interface to the point where their experience is "connect and play" and the leader has additional controls to run sessions efficiently.

The two use cases we run most often are a small choir for live-streamed religious services and a cross-continental band rehearsal cum recording session.

The choir is composed of 5-7 lay singers in their homes. At various times they've been as far apart as Baltimore, San Diego, and homes throughout the SF Bay area. The service is led by a Zoom host at their home who also runs the Telejam session. Zoom and Telejam are connected via an outboard mixer, also controlled by the host. The singers can monitor the service while not singing. When the choir sings, the output of Telejam is sent into Zoom through the host's connection.  The congregation is unaware of Telejam and hears the singers as though multiple Zoom participants are singing together. Of course, the Zoom audio is not in sync with the singers' images in the gallery.

The five-piece band (singer/ukulele, bass, keys, trumpet, drums) is located in Hawaii, Oahu, Brooklyn, Baltimore, and San Francisco. During weekly sessions we learn, arrange, rehearse, and record new tunes, then play back the recordings and discuss them. The interaction is like a live recording session. We use a Zoom channel running on the side for video, but most of the interaction is done using Telejam in both PLAY and TALK modes. After the session, any participant can download and remix everyone's tracks. The eventual goal is to produce an album.

We've also played with far-flung musicians to confirm our ability to work musically worldwide. The longest distance we regularly visit is between Bangalore and San Francisco, and from time to time we add players in Baltimore and Holland to that chain. Another international ensemble includes musical friends in Sweden, Holland, Baltimore, and San Francisco. We've also jammed with players in Israel, Italy, Switzerland, Japan, Mexico, and Brazil. Surprisingly, the greatest challenge has not been the quality of internet connections, or the distance and latency between players, but scheduling a session when everyone is awake at the same time.

## 7. A DIFFERENT DRUMMER

While implementation details are interesting, Telejam's  use of Web Audio and WebRTC is not unique to collaborative music applications.  The daisy chain design is novel, and it's worthwhile to note some of the differences between Telejam and other programs.

Upbeat and Acapella are popular programs that create multiplayer music videos. In many cases viewers believe that these performances were recorded live, but in reality they are made asynchronously. The participating players record by themselves to a backing track, send in their tracks,  and the final video is produced later. The experience is not a real time collaboration. Additionally, many choir directors bemoan the time it takes for everyone to submit their recordings before they can create their video. Telejam records tracks for all the participants while they are performing together. The director can record multiple takes in one session, with "all hands on deck" and remix the tracks as soon as the session is over.

Multiplayer jamming programs, like JackTrip, JamKazam, Jamulus, Sonobus, and others are all sensitive to the internet transmission latency between players. These programs attempt to squeeze latency to a minimum and provide setup parameters that each player can tweak. This places an extra burden on the players. The delays between any pair of players will rarely be the same, so synchronization is "loose" and becomes more of a problem with larger groups. Ultimately, all of these programs fail at long distances when the delays become unacceptable. Telejam does not ask players to perform latency adjustments and works over any distance because it disregards transmission latency.

Online digital audio workstations (DAWs), such as Soundtrap and Bandlab, permit asynchronous collaboration where multiple participants can record tracks by themselves and edit a mixing project while online together. They do not allow musicians to play and record together simultaneously. Telejam is not an audio workstation. It is a platform that lets musicians play together in real time. Telejam can record separate tracks for everyone while they are playing together. The tracks can be immediately imported into a DAW for further manipulation. Using Telejam along with an online or offline DAW can be a useful combination.

## 8. VARIETIES OF MUSICAL EXPERIENCE

Musicians are familiar with solitary solo performance. They practice alone, play along to recordings, produce tracks to be used later to assemble a finished product, and livestream themselves to an unheard and sometimes unseen audience.

Telejam is unusual because it occupies a space between a live in-person collaboration and a disembodied overdubbing session. It offers the ability to collaborate live, in the moment, with others.

Everyone playing in a daisy chain knows they are “in it together” and they are reacting in real time to each other. The spontaneity involved can feel like a live session.  In an interesting way, the activity on the chain is like the old MMO (music minus one) LPs, but the growing mix passed down the line is more like OMM - one more musician. The music each successive player receives becomes richer and richer.

Even though the first player has no incoming audio and plays alone, they can lead the group and influence downstream musical responses by setting the tempo, varying dynamics, articulation, and so on. All the players can feel like they are challenging or daring the players that come after them to follow along. The frisson can be exciting.

The word “latency” comes from the Latin “to be hidden, dormant, or concealed”, though in technical terms it is more commonly used to speak of time delays. Of course there is latency between players in Telejam, but since communication is only in one direction the participants are not aware of the time delay between them. In that sense, there is “no latency”; its existence is indeed hidden, and we can leave it behind.. literally!