Baylor School of Music Audio Collaboration Team

Report with Recommendations:
Using technology to improve the teaching and learning experience during the time of Covid-19 in the Baylor School of Music

August 10, 2020

This team was put together in the beginning of July to take one month to investigate a developing technology that enables audio collaboration in real-time over the Internet as well as to investigate audio enhancements of currently utilized video collaboration tools. This report is the result of this work. The group began with a survey of existing research, most notably from a team based at the New England Conservatory. The team then focused on leading software titles and technology to first learn them and then to test them out under a variety of scenarios for potential use with Baylor faculty and students. While much testing has been completed, more testing is needed and ongoing. The results are sufficient to share this report at this time.

The issue to address:

If the campus closes down to in-person instruction or if students elect to take their courses online, then real-time collaboration experiences will be minimal, if at all, within the School of Music using the current tools and teaching methods. The activities that would be unavailable under current instructional models for students at a distance include real-time interactive private lessons, rehearsals with accompanists, small ensemble rehearsals, sectionals, full ensembles and other time critical interactions. The systems that we have been investigating can be used to address most of these scenarios, but not all at this time.

For faculty and students in the School of Music, a shift to online learning removes an important and even essential component to the music education experience. Students might say:
“We need to make music with others to develop our musicianship. We need to learn alongside others. We need to create shared musical experiences. Indeed, we need to feed our souls with the collaborative experience of music and music making.” This document addresses the need for collaboration through the assistance of technology.

A second need beyond this high level collaborative environment is to improve the quality of the existing technology that the faculty and students utilized in the spring and that may have to be used again, including applications such as Zoom. We have confirmed that there is a free application to bring the audio quality up to the highest fidelity while enabling full two-way interaction with all participants. This, when paired with a faster video application will also improve the latency over a Zoom based connection. The result is still not quite to real-time quality, but it is closer to real-time than Zoom but with the bonus of high fidelity, two-way audio. This pairing reflects an improved communication experience.

One thing for sure of the need to teach and learn online at Baylor, “If you do what you did, then you get what you got.” We need to take in the experiences from the online teaching efforts of last spring, examine the tools and determine what improvements can be made. This document is the result of one team’s efforts at doing just that for the School of Music.

Potential actions to take for students and faculty

1. Do nothing different. This assumes that the tools that have been employed work with sufficient success that no other option or improvement is needed.
2. Take steps to improve the audio quality of the otherwise previously adopted video systems.
3. Adopt new audio technology to create a collaborative, high quality real-time teaching and learning experience.

A few words about technology

- Know the need, understand the options and then choose the most appropriate tool.
- Identify the problem first and then look for a solution second and not the other way around.
- When it comes to using technology to solve a problem it is essential to know which technology is appropriate to facilitate or solve the problem at hand.
- There is always a way.
- There is always a better way.
- As with any new technology used, there is going to be a ‘break in’ time in which operational skills are learned, often through trial, error and patience.
- Once a comfort level has been achieved, the technology blends into the background and the content becomes dominant.
- Technology is not easy and it will fail at times, but if the rewards outweigh the difficulties then the effort is worth it.

Available instructional options using technology:

1. Asynchronous (interaction separated by significant time)
2. Semi-synchronous (laggy communications)
3. Synchronous (lagless, or virtual real-time communications)

Asynchronous

Examples include email, web content, posted videos to watch at unspecified times, etc. This type of content delivery is not being discussed in this document as these sources are well known and are utilized as appropriate.

Semi-synchronous

This classification of video and audio content is marked by slight (but noticeable) delay in communication. While two-way audio is included either as full or half-duplex mode, the inherent delay in the transmission of the communication is sufficient to eliminate this category from providing a platform for full and natural interactive communication and collaborative music making. Participants can effectively use this level of quality for meetings, discussions and presentations but generally have to accept an environment that includes an amount of noticeable latency, also known as laggy communication.

Perhaps the best example of this type of communication is Zoom.

Zoom is a great option for many scenarios, but not for all types of needs. Features include multiple video squares of participants paired with average quality audio. The video and audio are generally in sync for each individual; however there is a delay of approximately 100ms (milliseconds) on top of transmission delays from each participant. This delay eliminates Zoom from use as an interactive, real-time music rehearsal tool.

Some negative features of using Zoom for music based applications include:

- The frequency range of the audio includes a rolloff in both the low and the upper range, thus limiting the possible sounds transmitted in the lowest frequencies in a sound source and limiting high frequencies that make up the upper overtones that are a part of musical sounds. Suffice it to say that the fidelity is lacking by our musical standards. Perhaps it is best to note that the audio response within Zoom is optimized for the more limited range of speech.
- The audio delivery is half-duplex, or one direction at a time and thus creating harmonies is not a possibility, nor are natural dialog interactions. Interactions can be compared to driving through an intersection in which all the traffic signals are blinking red. Everyone has to pause and wait for clearance to continue and it feels a little awkward. So too, can live discussions in a semi-synchronous environment.

Here are some of the positive uses of Zoom:

- Zoom is effective for presentations to large groups (more than 50) which supports a sharing of information by the presenter to the audience. This scenario is coupled with limited or no functional interaction as there are just too many participants to allow for a free exchange of communication from all participants.
- Zoom is also effective for interactive meetings of perhaps less than 50 participants in which discussions could be had. The higher the number of participants makes the free flowing interactions more of a challenge.
- Zoom excels at meetings with a small number of participants, but again, the method of interaction includes laggy communications.
- Zoom is currently used for private lessons, which does work to a degree with the major exception of not allowing for real-time interactions such as rehearsing together or accompanying the student. It also does not provide an accurate representation of the sounds of the musician.

Recommendations for Semi-Synchronous connections

What we would like to suggest are ways to improve the audio quality, its format (full duplex for natural interactivity), and a way to lower the latency of the video to bring the semi-synchronous nature of this closer to real-time. This will make a difference in the overall experience and will be relatively simple to employ.

Audio Software to use

There does not seem to be any (free) video based application that also includes a high fidelity, full-duplex audio option with low latency. With this in mind, the suggestion is to use a video application and have everyone mute the mic audio (as you will not use it). For the audio signal, use Cleanfeed. The free version of Cleanfeed is quite sufficient for our use but there is a pay per month option for those that want even more. The audio clarity and functionality is striking in comparison to the current audio of Zoom and other similar applications. This is not a synchronous solution but is a better semi-synchronous option than many others. A much more in depth study of the Cleanfeed option is included later in this document.

Video Software to use

One could use Zoom as the video software, but remember that its performance comes with a relatively big delay or latency. We are seeking a faster video solution, one that will push more towards real-time video. Currently, given most of the use scenarios and that we are using the commodity Internet, real-time video is not a possibility; however, we can use a video package which has some adjustments to enable quicker throughput and lower latency. This is going to be accomplished by adjusting the video quality lower so that there is less data information to transmit for the video signal. Zoom does not have this feature end to end but Jitsi Meet does. Jitsi is also free. This web based application is found at https://meet.jit.si

Choosing to make this change in approach to video communications for student to faculty interactions is relatively easy and the only cost is in making a couple of settings adjustments and in utilizing two applications to handle the connection instead of one. This should be an easy change to make since the result for musicians is a noticeable improvement in the quality of the session.

Synchronous

This is the category that has been a main focus of this investigation and recommendation document as this presents an option to potentially reach a virtual real-time experience. It involves the enabling of real-time music and arts collaboration through the use of technology to overcome a physical distance with a perceived, in-person audio experience in real-time.

What is real-time?

In reality, real-time is non-existent. Air is a resistor to sound. From the instant a sound leaves a sound source and travels through the air, it is going to arrive with a delay. As an illustration, it is said that you can tell how far away a lightning bolt struck by counting the seconds before you hear the sound of thunder. Light travels much faster than sound and both travel at generally consistent speeds, so the delay time can be converted into distance.

Rather than express the delay time in seconds, we need to quantify it in milliseconds as we seek to collaborate as close in time together as possible to simulate being present in the same space with those we are connecting with.

1 second equals 1000 milliseconds. Therefore, 1 millisecond is $1/1000^{th}$ of a second. A workable formula for how fast sound travels is 1.125 feet per millisecond. If you were standing 1.125 feet from a sound source, then it took 1 millisecond (1ms) for that sound to travel to your ears. Even when in close proximity to others making music, there is a delay and yet our brain interprets that the sounds are 'together' and in 'real-time'. Clearly, the determination of real-time is one of perception. Just as some musicians have the ability to discern intonation and 'in tuneness' at more accurate levels of frequency as measured in cents, humans also have varied levels of discernment of time as being in 'real-time' or as detecting a delayed sound as measured in milliseconds.

For our purposes and understanding, the range of delay which most people discern as being in 'real-time' is between 1-@35ms with a goal of achieving around 25ms connections. To express this in another way, 35ms equals 3.5% of one second or about the time of a single 128th note in a measure that has a tempo of quarter=60bpm. A connection that is nearer to 25ms would be in the range of a dotted 256th note at that tempo. This is the level of quickness and togetherness that is the goal for the highest level of real-time communication. As an example at the human extremes of perception, one could surmise that a professional percussionist would have a lower threshold of perception (lower than 25ms) and a non-musician would have a higher threshold relative to each other as they determine whether multiple sounds were occurring at 'the same time'.

As the sounds become perceived as out of time sync, the brain starts to compensate and will adjust as much as possible to tolerate the imprecise match in time up to a point. That point will be crossed well under 100ms difference in time and that millisecond threshold is different for each individual. Still, we can target our efforts at achieving delay, or latency under 35ms.

Following this thought, the acceptable threshold of milliseconds may be influenced by the type of musical sound source. For example, a legato start to two string instruments sounding together or two voices may influence and relax the perception of the threshold of what is real-time as compared to two handbells or snare drums playing at the same time. This is something that only experience will determine if it is factor for our perception of timing. Therefore, an important consideration is that the instrumentation involved may point to greater perceived success at slightly elevated millisecond numbers.

In summary to this point, in music or in communication from person to person there is no such thing as absolute real-time but instead, we are trying to work to achieve <u>perceived</u> real-time.

<u>Recommendations for Synchronous connections</u>

This ability involves leading edge technology that is not yet to a point of automation or use without user involvement. In other words, there is not a plug and play type of package to participate in a synchronous connection….today. Synchronous connections are a challenge because they require the overcoming of multiple and unique variables in both user hardware and network connectivity for each participant. This can be accomplished through individual tweaking of attributes involved in a connection if the software allows. Overall, this is pushing the boundaries of what is possible over the Internet at this time and yet, because it is potentially possible and our situation towards a musical collaboration need is urgent if the circumstances fall towards learning at a distance, this becomes a more important consideration. Essentially, the methodology needed to trim off milliseconds of time in audio connections over the Internet is accomplished by adjusting what can be adjusted. Such adjustments would not speed up the transmission, but rather would speed up the processing of audio data at the endpoints and within the routers involved. The adjustments to be made include the size of the data packets created and the amount of buffering of the sent and received packets.

Audio Software to use

There are several entries into this field with members of this team having experienced and tested at least four of these options. They all attempt to lower the latency numbers with some able to accomplish this more than others. In addition, each comes with its own set of technical adjustments available with some more complex than others. The need is to find a blend of performance with as little technical skills required to produce the best results. With these desires in mind we have settled on Soundjack as enabling success balanced with the need for new learning to take place. Soundjack is free to use. Much more about Soundjack is included later in this document.

Will this have a chance at success?

A key point to make at this time is that success with Soundjack or any of the synchronous options will depend on many conditions or factors including the following:

- the processing power of a computer,
- the use of an external mic (USB or a higher quality mic connected to an external audio interface),
- the use of headphones,
- the use of an Ethernet connection to a home users router or to an Ethernet drop in the faculty members office,
- an Internet connection with sufficient bandwidth,
- a network connection to and from the other participant/s through the Internet that does not hinder a clean connection that includes as few hops (intermediary routers) as possible, and finally,
- the patience to learn how to operate in this environment.

Despite this daunting list, it does work, but it is most important to understand that not everyone will likely meet these conditions. Some students or faculty may not have the basic technology, infrastructure, network connectivity or patience to make it work. But for those who do, making music together while separated by distance can be a real thing.

Solutions in Detail

Details on Cleanfeed

Cleanfeed is a software solution that allows for relatively high-fidelity audio transmission over the Internet and has shown to be an effective tool for virtual music instruction.

How it works:

The software is activated through the use of an Internet browser on a computer; no extra downloads are required. Only Chrome and Brave browsers are supported. The host user must first create a free account at cleanfeed.net. (There is a paid tier, but its features do not fundamentally improve the usefulness of the service for music instruction.) Thereafter, the user logs in to start a session. Once logged in, the host can invite participants to join a session by typing in an email address on an invite form. Invited users do not need to have an account. They will receive the email invitation and click a link to be taken to the session page. The only additional equipment needed is an external mic connected to the computer (or to an external audio interface) and a set of headphones.

The interface is very simple and easy to use. Essentially, there are only two settings for the host and one for each guest. The host must select the global audio quality ("music optimized") and the microphone source. The guest must select only the microphone source. All participants can control their mic with a mute or unmute button. Headphones are necessary because, unlike Zoom and similar conferencing platforms, Cleanfeed does not employ echo cancellation or noise reduction. This improves sound quality but requires isolating the local audio source from audio playback to prevent system echos.

Dr. Brian Marks writes:

"I have used Cleanfeed to teach a number of virtual piano lessons over the course of the summer and am very pleased with the results. The sound quality is definitely far superior to using Zoom or Facetime with a smartphone or a computer using the built-in mic and speaker. While it is possible to improve the audio quality of Zoom by adjusting a few advanced audio settings and using an external mic, given the same conditions, Cleanfeed still produces a superior result. Compared to virtual lessons I gave in the spring using Zoom and Facetime, with the Cleanfeed lessons I was able to hear a much greater range of dynamic and tonal differences in my students' playing. Although not as clear as a face-to-face lesson, Cleanfeed allowed for a highly productive virtual lesson setting.

I experienced a few stumbling blocks in my summer lessons that will be useful to keep in mind. First, Internet speed and microphone quality do matter. One student I worked with had relatively slow home internet and a fairly cheap external mic. The sound quality from this student was of noticeably lower fidelity than my other student, who had a higher quality mic and faster home Internet. Yet, even in the case of the student with the lesser equipment, the audio was good enough for us to cover fairly sophisticated musical elements.

Second, different operating systems may have greater or lesser ease of use. Both of the students I worked with used Windows laptops and each had some initial difficulty getting their computer to recognize the headphones they were using. One student plugged his headphones into the audio jack of his laptop, while the other used Bluetooth headphones. Once they solved their connection problems, everything worked smoothly thereafter. Using a Mac laptop, I experienced absolutely no problem with my headphones being recognized.

In summary, I think Cleanfeed can be an extremely useful tool for virtual music instruction and I would encourage every Baylor music faculty to become familiar with it. I would expect that once they do, they will use it extensively for virtual music instruction."

There are a few considerations that need to be kept in mind when using Cleanfeed:

1) This is an audio-only platform. For video, participants can simultaneously use Zoom, Facetime, Jitsi Meet or another video conferencing tool with the sound muted while utilizing the audio feed from Cleanfeed.

2) Practically speaking, Cleanfeed must be used on a computer. Technically, one can host a Cleanfeed session on an Android device using Chrome, and it is possible to use an iOS device as a guest. However, the need to use an external mic and headphones for music audio quality really limits the hardware platform to being a computer.

3) A quality USB mic improves the overall experience considerably as would a higher quality mic plugged into an external audio interface.

4) Quality headphones are encouraged, but are less important than the quality of the mic. One must avoid using the built in or external speakers to hear the remote participant/s so as to avoid any audio loopback sounds.

5) Cleanfeed does not solve latency problems, so it can't be used for synchronous rehearsal or even for the teacher to play along with the student.

6) The host can invite multiple participants for a music based session, but the latency issues mean that only one person can make music at a time. Cleanfeed would be an enhanced sound source for standard video based meetings and presentations.

Useful links:

Main site: cleanfeed.net
Cleanfeed FAQ: https://cleanfeed.net/knowledge/faq

Details on Soundjack

Soundjack originated as the work of a Ph.D. candidate over 11 years ago in Germany and has been under continuous development by its creator. Soundjack has always been free to use which is in keeping with the software creator's desire to do something good for the larger community. It was not until the pandemic that this started receiving extensive interest because of the obvious implications of a successful implementation.

Soundjack is not a traditional application. Its program, that is initiated on a local computer, runs in the background and only has the option of quitting. The rest of the action takes place on the Soundjack web interface. Adjustments made there are then enacted on one's computer via the background connection to the sub-routine running.

The process to connect into a session includes first starting the background application and then accessing the Soundjack web page. Once logged in, the user can do several things such as view the Tech Tutorials but most importantly enter the Stage. This is the web page where all of those online are listed and can be 'called'. It is also the single page that includes the various settings that are in place, including the three main settings that can be adjusted to nuance the quality and the latency of a connection. One does not have to know or understand everything shown with the exception of how to manipulate those three settings to optimize a connection. We feel that this is a very doable thing and can be taught to our students and faculty. With any experience, users will become adept in doing this and can really focus on the music making.

That said, like any new technology or program, there is a learning curve and a level of confidence that comes from experience. This does not seem nearly as steep of a curve as other programs that students and faculty use. One of the side potential benefits is that for some, they would work to obtain the optional enhanced tools which happen to be of direct benefit to musicians. This includes a quality mic or two and an audio interface. If music students do not already have such equipment, it is not out of the realm of appropriateness for them to obtain these tools for other uses as well.

How it works:

Soundjack operates its connections as peer to peer. This means that there is not an intermediary server involved where the signals are blended, rebroadcast or reflected. This is an important concept and one that comes into play with respect to making connections within the Baylor WAN. Traffic that originates on campus to another on campus user all stays within the on campus network. Because of this, the connection speeds are extremely fast and definitely fall into the category of real-time. Again, this is a concept that we should be able to take advantage of to solve some of our space issues.

When a call is made, a connection from one computer to the other is established. From this point, the adjustment capabilities are employed by each user in a way that will compensate for the networking issues of the data path as well as to compensate for computer endpoints that have a variety of processing power. Poor computers mean that adjustments to the buffering must be made to overcome the delay in converting the audio data into audio sound. Buffering adds milliseconds of time. A poor network connection or one in which there are an abundance of intermediary routers also can be overcome to a degree with the same adjustments in Soundjack. Again, added buffering adds to the total millisecond time of the connection. The goal becomes to make all adjustments in a way that is on the edge of audio stability while minimizing the latency. This ability to adjust is what enables the lower latencies and is what is not present in other mainstream applications such as the audio portion of Zoom. This ability to adjust is the 'secret sauce' in this effort to attain the lowest latency numbers. Once the parameters are set, only an occasional readjustment may be needed or desired in response to fluctuations in the network connectivity in a session.

Can video be added to a Soundjack session?

Yes, but not with the same low latency qualities of the audio.

If we are achieving real-time audio collaboration with Soundjack for instruction or small group rehearsals, then an important addition would be the ability to pair the audio with a video feed. This would be utilized by the faculty instructor in the case of lessons or by a director in the case of a small ensemble for instructional purposes. The amount of data for a video feed is more than there is for an audio feed. Because of this, it is currently not a feasible expectation to achieve real-time video delivery at the same level that we are achieving audio success.

That said, there are several options to incorporate a semi-synchronous video element as a reference tool to pair with the low latency audio feed of Soundjack.

As is the case with the relationship between high fidelity audio and the amount of data generated by it as compared to lower fidelity, so too is the case with video. The amount of data associated with a 4K stream is tremendous as compared to a low-res, black and white video. Less data to send and receive (whether it is audio or video data) equates to faster processing and this translates to lower latency. Therefore, if we want to utilize a lower latency video feed to accompany and provide a closer sync with the low latency audio, it will need to be a lower resolution quality of video.

When adding a video application to pair with Soundjack, one would have all participants mute the audio portion of the video application in order to utilize just the audio from Soundjack.

Here are some options:

> Zoom: The current version of Zoom, for example has several settings to change in its Video Settings tab including unchecking Enable HD and on the Advanced page, to uncheck some of the settings related to enabling hardware acceleration. These changes may reduce some local machine processing requirements for video which may help the focus on the processing for audio and its low latency goal. Testing with these changes

has not been done to determine its effects on the system. Because of the design of the Zoom model, the incoming video latency may or may not be affected by these local changes. Zoom is free.

JitsiMeet: The use of Jitsi has been tested and shown to have a lower video latency in general, especially when used with two participants. This is because Jitsi operates as a peer to peer design with two participants but converts to a hub and spoke, or intermediary server based design for three or more participants. In addition, there are settings to change the video quality down from the default of High Definition to Low Res and this has an effect on the video latency. Jitsi will provide a better experience for participants due to its closer sync of its video to the Soundjack audio. The look and feel of Jitsi is similar to Zoom, although not as developed. Jitsi is free.

Soundjack: There is a video component within Soundjack that potentially can become the best option; however, this feature is currently considered in beta and can be unstable for connections. It has been said that this feature will be addressed by the developer in the future. Until then, its use is partially recommended. The key aspects that make this a desirable solution begin with the video quality being limited to low res to start with. There are additional settings to lower the quality even further as well as to enable black and white video only. Again, the lower the video fidelity, the less quantity of data to process and the lower the latency will be. Turning on one's video sends it to the other participants while not all participants would need to send video back. For our use, this option is important in that we could have only the director send video to the participants in the rehearsal.

Other: There are undoubtedly other video conference based systems available. Look for those with low latency to start with and those that allow for the reduction in video quality. As with all of these, one would use the video and mute that system's audio when pairing with Soundjack or Cleanfeed.

Current recommendation for pairing video with Soundjack or Cleanfeed:

1. Jitsi Meet with audio on mute
2. Zoom with audio on mute
3. Soundjack

Useful links:

Main site: https://www.soundjack.eu/
Soundjack Tutorials: https://www.soundjack.eu/index.php/howto

FINAL RECOMMENDATIONS

1. Audio improvements to Zoom style activities:
   When advantageous to the activity, that the concept of a video based session be structured to include an audio connection of all parties using Cleanfeed to be paired with the video only portion of the video application.

2. Audio interactions in real-time:
   That there be a phased-in approach to the implementation of the low latency, real-time solution, Soundjack through the semester with an initial focus on gaining expertise through experience with one to one scenarios such as private lesson instruction or student rehearsal with an accompanist. With experience, then begin to expand the number of participants to include small groups working together such as a student with an accompanist with an instructor, and groups of students up to quartets and quintets. A slightly different phase in may be appropriate for other studio or ensemble sectional needs.

3. Training:
   The need for training is obvious but the urgency has to be tied to the anticipation of the need. Cleanfeed will be easier and quicker to learn and implement than Soundjack.
   - Cleanfeed:
     - Training for this application should be fairly easy and quick. Ongoing support should be minimal.
   - Soundjack:
     - Provide an overview to all potential users early on to provide awareness of this opportunity.
     - Because there is an initial assessment of current hardware and network abilities for each participant with the potential delay of obtaining more appropriate hardware, efforts should begin as soon as possible with early volunteer participants. In addition, it should be expected that there will be a learning curve associated with this higher level technology and that testing will establish confidence and of course, will take some time. These early volunteers would become peer assistants to new participants as they join in on the use of Soundjack at later dates.
   - A determination of the training methods and personnel resources should be made quickly to prepare for upcoming presentations and trainings.

4. Equipment and Infrastructure:
    - Equipment for Cleanfeed
        - To improve audio means to improve the audio capture through the use of a USB microphone at the minimum.
        - If the teaching or participation situation warrants, then an upgrade to an external audio interface paired with multiple XLR connected mics with mic stand/s be sought. The selection of these mics would be made based on the anticipated uses by the participants.
        - Headphones will be likely required. Wired or wireless should be fine.
    - Equipment for Soundjack
        - To improve audio means to improve the audio capture through the use of a USB microphone at the minimum.
        - Depending on the processing power of the users' computer and the added latency to a connection from using its built in sound card, a separate audio interface with one or more XLR connected mics with mic stand/s may be needed to optimize both the audio quality and to trim milliseconds off of the overall latency of a connection.
        - We need to further test the dedicated computer substitute devices (Raspberry Pi based Fast Music Box) for potential deployment in selected practice rooms or to be recommended to students who only lack an appropriate computer.
    - Infrastructure to support the use of Soundjack:
        - One of the likely widespread uses of Soundjack will be the inclusion of an accompanist with a student either as they are preparing for a recital or as part of a private lesson. Not many of our student accompanists own their own pianos and therefore make extensive use of SoM quality pianos in certain practice rooms and rehearsal spaces.
        - To this point, and knowing that Soundjack operates best if in an Ethernet wired connection, we recommend that steps be taken to provide an Ethernet drop in each of these critical piano spaces.
            - Clearance will need to be made with IT as to the connection of various computers to these Ethernet ports.
            - Knowing that this approval will be challenging, we should consider further exploration of the use of the dedicated Raspberry Pi based Fast Music Boxes to provide a dedicated, purposed and approved substitute for a computer in these rooms.
        - We should investigate the possibility of adding a Soundjack Mix Server within the Baylor WAN to determine if it could assist in lowering the latencies for those who otherwise would be on the fringe of such levels of connectivity.

- Further enhancements in the campus wireless network, AirBear should be investigated and sought to see if the resulting latencies inherent in wireless connectivity can be reduced to an acceptable point while making connections within AirBear.

Respectfully submitted by the Audio Collaboration Team
August 3rd, 2020.

Team Members:

- **Mr. Bob Avant, Team Coordinator**
- **Dr. Mark Diamond**
- **Dr. Jamie Van Eyck**
- **Dr. Ben Johansen**
- **Dr. Brian Marks**
- **Mr. Alex Parker**
- **Dr. Amy Petrongelli**
- **Dr. Lauren Weber (Theatre Dept)**

Addendum 1:

Understanding data transmission over a network:

In theory, data travels over the Internet at a similar speed or capacity consistently. In practice, it never does. Its actual delivery speed is always influenced by the amount of data congestion on a given transport media type. It is also affected by the processing power of each router and the quantity of routers that it encounters in the path from point A to point B. Each hop (router processing point) adds to the millisecond total of the overall time for the data transmission. Since there is not a stipulation for a deployment of a router per fixed distance in mileage, there can be a wide variety of total routers involved over any given physical distance. However, physical distance can be an indicator or predictor in the potential for low latency numbers to be present in a connection. The current thinking is that low latency using one of the best communication tools presented herein can be made possible perhaps up to around 500 physical miles of separation. It is more likely to provide successful connections up to 200 miles or less in distance. One last characteristic that strongly influences the potential for very low latencies is the proximity of a LAN to the backbone of the Internet. Most home connections are many hops away from this backbone and will be a negative contributor to success, but not render it impossible. The Baylor WAN is positioned much closer to the backbone and this is a positive influence on connections that would be made from outside of the network to within the Baylor wired network. (Wireless networks and networking add milliseconds unnecessarily and hinder success). Knowing this, Baylor faculty could initiate connections from their SoM offices to students who are just off campus or even from different cities with an enhanced chance of success.

The quality and consistency of an audio connection over a network is also enabled or hindered by the processing power of the computers used by the participants as they convert the audio to data and also the reverse. The data packets that are sent over the Internet must be created, sent, received and processed within each user's computer. This takes a computer that has a decent capacity for processing. Older computers may or may not be of sufficient quality for this purpose. One general indicator is that the processor has at least 4 cores. This does not exclude those with only 2, but it does indicate the increased likelihood of success. The communication software system that we are suggesting has adjustments available to optimize a connection from unique computers and their networks. There is not a one-setting-for-all because what we are doing is trying to shave a few milliseconds here and there to get the connected users to connect at the lowest possible millisecond rates given their unique collection of variables. Specifically, the adjustments are in how the data packets are assembled and in the quantity of the buffering of data packets being sent/received to allow for sufficient processing time within the computers involved over the network. Because there is more capability for adjustment of settings, this can translate into the perception of complexity. This is the tradeoff that provides us the best chance to achieve low latencies.

This begins to identify a likely condition with our students, that there will be some who would be classified as 'technology have-nots' based on a hardware issue or a connectivity issue. To what extent this may be the case is unknown and would only become known by pre-testing each potential user individually. Even within the team members that worked on this research were examples of users who, because of their combination of hardware and networking would have a challenge to participate with the desired lowest latency numbers. There may be some opportunities for technical or financial assistance from the University that some of our students could take advantage of if needed to acquire stronger computer connections.

To continue with the topic of the transmission over a network, be that a LAN, WAN or the greater Internet, there are two types or connection topologies being made in these audio systems. One is described as a 'hub and spoke' design and the other is called 'peer to peer'. Just by its design the peer to peer type of connection will be faster than the hub and spoke method. For hub and spoke design, each participant connects to the hub (an additional server where signals are mixed together) and on to the other participant/s. With the peer to peer model, the connection is made directly between the participants. If the participants are in the same network (such as the Baylor WAN), then the connection is made through a path that includes the first common connection router. In this example, the main Baylor router would be as far as the signal would need to go in order to turn around and make the connection back inside its own network. This should and does generate the lowest latencies of all and are well within the range of what is considered as 'real-time'. This would mean that wired 'on campus to on campus' connections should be very successful. The team has tested this and found this to be the case.

One aspect of making low latency connections is that utilizing a wireless connection is not supportive. Wireless traffic is inherently unstable and connection speeds vary greatly every second. Normally, the system overcomes this variance by the addition of buffering. This adds to the latency which we are trying to keep at a low and consistent rate. One potential concept is that if a student is on the Baylor Wireless network and is making a connection within the Baylor WAN (a separate network), then the cumulative latency may still be within an acceptable level because of the speed of the WAN in general. Because we have seen a variety of test results, this is something that we need to continue to test but it does have a potential to help some of our connections with students on campus.

Making these quick connections solely within the campus network thus carries a great deal of potential meaning for our situation and needs to be fully thought through as to how to exploit this condition. This may be a way to provide on campus based collaboration without the risks of having the participants in the same room. This could relieve the demand, use and scheduling of limited rehearsal space and all of the mitigation associated.

ADDENDUM 2:
Selected Commentaries by members of the Team:

- Students and collaborative pianists are major benefactors of the implementation of this technology
- Soundjack is certainly an excellent solution here if the technology/hardware allows it
- Soundjack could work in a synchronous applied setting, but I see more value with less technological woes using Cleanfeed/video platform of your choice
- Sound quality and clarity is far more important than latency issues
- I am amazed at the improvements made since March technologically and we should keep our eyes/ears peeled because I believe changes will continue to come in the near future.
- We want students (primarily) to be able to collaborate in conditions where face-to-face contact is limited by space constraints (insufficient number of sufficiently large rehearsal spaces that allow for social distancing) or lockdowns.
- Low-latency virtual connections could significantly increase useful spaces where (virtual) collaboration could take place.
- These changes offer more natural conversations
- Thinking long term, even in 'normal times' this ability will become more accessible to more musicians in a way that will open new doors to collaboration even to the point of

this being considered normal. Our students would benefit from and perhaps enjoy the experience of this first generation of such collaboration in the digital age.

- The benefit isn't just low-latency, but also the ability for both parties to speak/sing/play/make noise at the same time.
- Low latency allows students to work on their skills to make music together in “real time” which is central to training students on collaborative singing.
- Certainly, this could be a critical enhancement to students who are in rehearsal with an accompanist as they prepare for personal recitals. Pianists would be in one of the SoM practice rooms or other space with a piano and the soloist would be anywhere else, including on campus, in Waco or even a different city.
- One need is to equip rehearsal rooms with either high-speed wifi (desirable for the flexibility of devices that can connect) or wired Ethernet (potentially faster and cheaper to install, but less flexible in the hardware that can connect to it).
- Depending on the individual, there can be a major technological learning curve
- Soundjack is tricky, and getting over the initial learning curve is daunting. Just as when dealing with any technology, one must be patient and devote time to learning the system.
- There must be solid internet connections on both sides of any low-latency connections.
- There will be obstacles of hardware, connectivity and knowledge. We can be involved to actively address some but not all of these issues.
- There will be some who will excel in this environment and some who will not.
- There are potential hardware upgrade costs for those who do not have the needed minimal equipment.
- There certainly are benefits to low-latency connections, but I believe CleanFeed would be the best option in most cases for clearer sound, as Soundjack requires a larger time investment up front. If latency is not a concern in individual lessons, even something as simple as Zoom with a good mic set up (either USB or with an interface) would also be a workable option.
- I think that 2-person collaborative scenarios, most likely soloist plus piano accompanist, are the most immediately beneficial and easily deployable scenario using Soundjack. Once a baseline of experience and problem solving has been established for this scenario, then situations involving more participants may be explored.
- Low-latency virtual collaborative technology is crucial for dealing with the current crisis, but it will not go away once the crisis abates. Once established, virtual collaboration will continue to be a part of the music-making landscape in the future. Baylor would do well to become proficient in its use in order to stay competitive as a music institution in the future.

ADDENDUM 3:
Hardware Recommendations for Baylor Students and Faculty

Baylor School of Music Hardware Recommendations
for Online Lessons and Audio Collaboration
August 3, 2020

The following recommendations will help maximize the online lesson experience or online audio collaboration experience. The "Entry Level," or basic, recommendations have been made while keeping costs as low as possible. The "Enhanced Level," is for those who would seek the higher level effort at real-time audio collaboration or for those who would be teaching online.

- If you already have an equipment setup that is satisfactory for both the student and the instructor, there is no need or suggestion to change or purchase additional equipment.

- We recommend use of a laptop or desktop for online lessons for both the student and the instructor. Smartphones and tablets will not allow the use of an external microphone and headphone combination, and will not allow the use of the recommended audio and video platforms.

- Hardwire the computer to the home router or to a network Ethernet connection drop using an Ethernet cable. This avoids the latency problems associated with wireless connectivity.

- We recommend that an external microphone be used for online lessons, as that is the factor that most significantly affects the quality of the music and interactions.

- Close all unneeded browser windows and all unneeded applications before the session including any processor intensive applications that may be running in the background.

- We recommend the following platforms in tandem for online lessons in particular.
    - Audio:
        - Cleanfeed (www.cleanfeed.net) is a free web-based audio platform that is a remarkable improvement over Zoom audio.
    - Video:
        - Zoom or Jitsi Meet (meet.jit.si), both are at no cost.
        - Mute the Jitsi Meet or Zoom audio when using either in conjunction with Cleanfeed. Jitsi will likely offer lower latency video that will better match with the Cleanfeed audio. Please be aware of potential security issues with Jitsi Meet and enable a password access to your video sessions.

- Both Cleanfeed and Jitsi Meet require use of the Chrome browser. Also know that Baylor Information Technology does not offer support for the use of the Jitsi platform.
- The Cleanfeed and Jitsi Meet platforms are also recommended by the New England Conservatory Voice and Sound Analysis Lab and other institutions.

Hardware Recommendations

Entry Level

This level represents a basic setup that enhances sound quality. Obtaining an external microphone is the top priority. Headphones should also be used. Wireless earbuds that you might already own may suffice for lessons but possibly not for audio collaboration. When it comes to price, the cost of an item speaks to its quality, durability and long term usability. Choose according to your needs, budget and future intent.

- External USB microphone (price range: $50-$150)
- Headphones: earbuds likely ok, wired is preferred. Open-back headphones may be desirable in some scenarios with closed-back being more likely to be chosen for other scenarios including use with recording applications (price ranges: earbuds $20 and up; open-back $100 or up; closed-back $50 and up)
- If recording capability is needed, then consider Garage Band or Audacity software (both free).

Enhanced Level

This level represents a more sophisticated setup that could serve as the core of a home recording setup or for a faculty studio setup. Using the equipment recommended at this level requires more technical expertise and is not surprisingly more expensive. A laptop or desktop computer is at the center of this setup with a preferred 4-core processor for use with audio collaboration, although 2-cores will work for most other applications. This level also requires both an audio interface and standard XLR microphones, cables, and stands. Because of the use of an audio interface, a wider variety of microphones can be utilized.

- Audio interface (PreSonus Audiobox96 or Focusrite Scarlett 2i2 or equivalent at the minimum), (price range: $100 and up)
  - Note 1: make sure the interface purchased is compatible with your current operating system and USB interface.
  - Note 2: Windows users may want to avoid the specific model of the Scarlett 2i2 if they are going to be using Soundjack for audio collaboration as there is currently a known hardware conflict.

- One or two standard XLR connection microphones, cables, and stands, depending on the intended use. (typical price for two microphones, cables, and stands: $285)
    - Note 1: There are different categories of microphones including dynamic and condenser with various pickup patterns available. In most cases, cardiod patterns are normal and suggested. Your selection should be based on your need as a vocalist, instrumentalist, or a combination of the two.
- Headphones: wired if possible, earbuds likely ok but wired is preferred. Open-back headphones may be desirable in some scenarios with closed-back being more likely to be chosen for other scenarios including use with recording applications (price ranges: earbuds $20 and up; open-back $100 or up; closed-back $50 and up)
- If recording is a part of lessons: Garage Band (free), Audacity (free), Logic Pro ($199), or Ableton Live ($269)
