Audio Engineering Society

# Conference Paper

Presented at the 2021 AES International Conference on
Audio Education
2021 July 22 – 24, Online



# Networked Music Performance: Developing Soundjack and the Fastmusic Box During the Coronavirus Pandemic

Christian Hoene[1], Ian Howell[2], and Alexander Carôt[3]

[1]*Symonics GmbH, Germany*
[2]*New England Conservatory of Music, Boston, MA, USA*
[3]*Hochschule Anhalt, Germany*

Correspondence should be addressed to Christian Hoene (`christian.hoene@symonics.com`)

### ABSTRACT

Networked music collaborations are technically demanding because of tight latency requirements. Typically an acoustic maximum one-way delay of 25 ms needs to be achieved if two musicians want to play in sync. We have developed the Fastmusic Box (FMB), a system that automatically analyzes the given system and network setup to identify all issues which might hinder a low delay transmission of audio. In this publication, we describe the algorithms implemented into the FMB. Since social distancing was imposed early in 2020, hundreds of music students and teachers have used the FMB to continue singing, playing, and rehearsing music together. This paper also reports on their experiences with the FMB.

## 1 Introduction

The SARS-CoV-2 pandemic of 2020 and 2021 presented unprecedented challenges to musicians, music students, and educators [1]. The risk of viral transmission was thought to be especially high for singers as they both produce and expel droplets and aerosols in larger quantities than their instrumentalist colleagues. Early reports pointed to choir rehearsals as particularly risky events, likely to cause an outbreak. For example, on March 10, 2020, a Skagit Valley Chorale (SVC) rehearsal resulted in the infection of 53 of 61 members. Two singers died [2]. Given insufficient data to confidently rule out singing as a significant source of viral transmission, and enough evidence to highlight the risks, singing with others in enclosed spaces was largely proscribed by governments, academic institutions, and businesses for the duration of the pandemic [3]. In settings where singing indoors was permitted, restrictions such as regular testing, masking, social distancing of up to sixteen feet, significantly shortened duration of singing, room restrictions following use, and air ventilation were frequently imposed. According to studies, these restrictions could mitigate the risk of transmission, though in a manner detrimental to the normal act of singing [4]. Additionally, travel to an in person lesson, rehearsal, or performance may impose the transmission risks inherent to taking public transport and moving through public areas.

Due to the health risks and logistical challenges, many teachers and students looked for alternative ways to continue their artistic work, rehearsals, and teaching. The only way to eliminate, rather than mitigate by degrees, the risk of transmission is to be separated in different rooms with separate air masses. This led many to explore teleconferencing solutions. However, teleconferencing solutions such as Zoom, Teams, or Skype are insufficient for musicians' needs as the audio quality is not optimized for music and the latency too high to allow for real-time interaction [5]. Off-the-shelf conferencing and telephony solutions have a one-way acoustic latency — measured from the mouth of the speaker to the ears of the listener — on the order of hundreds of milliseconds. For speech, this is not an issue as the human ear tends to tolerate delays up to 400 ms. For musicians, this would be the real-world equivalent of collaborating while standing further apart than the opposite end zones of an American Football field.

This, however, can be overcome within a constrained area such as a campus that has a managed communication network. Here it is possible to set up a system to transmit video and audio signals between rooms. Technically, this can be achieved via traditional analog wiring or digital communication protocols such as AES67 or Dante for audio, and SMPTE 2022 for video.

The third technical options are Networked Music Performance technologies (NMP), which can also be considered a teleconferencing solution optimized for low latency [6] audio and/or video. For such a system to work, every step in the process of transmission must be optimized to avoid the introduction of unnecessary latency. Once transmission delay is introduced, it cannot be compensated for. For example, the physical constraints acting on these systems, especially the speed of light, give hard limits on the maximum achievable communication distance. In an optical fiber, within 25 ms a maximum of approximately 5,000 km can be traveled. Eventually, the final latency depends on the physical signal propagation, the speed of light, the transmission latency according to the available bandwidth in each switch or router, and the actual transport technology. Given these limitations, a technical base setup aiming for the lowest possible latency is required.

In 2019, one of the authors organized a remote music session via 5G, including locations in London, Birmingham, and Bristol. Jamie Cullum gave a lecture/performance in the famous Guildhall, London, where he performed songs with amateur musicians on the other sites [7]. Soundjack [8] was used as the NMP technology, which transmitted audio and video via a 5G network. Overall, the event was well perceived by the musicians and 150 guests, showing that NMP is technically feasible to conduct remote music lectures. It should be noted, however, that this event was only possible with technical experts present before and during the show at all sites. Also, the charity "Music for All" [9] did great work in organizing the event. If teachers and students want to play together on a daily basis, it is unreasonable to expect them to possess the necessary technical knowledge to execute such an event. Instead, the NMP system must work as smoothly as well-known and established conferencing tools.

Indeed, bringing NMP to the largely non tech-savvy community of music performers, educators, and students directly affected by pandemic restrictions is what we consider our main technical challenge. In order to overcome this, we developed the Fastmusic Box, a technology that helps musicians simplify the challenges of NMP. Additionally, we published guides, tutorials, and advanced workflows, gave international online workshops, and developed curricula. Furthermore, we supported the NMP of hundreds of musicians during the pandemic via support channels and through social media.

What follows is first an overview of the Soundjack streaming architecture which consists of a core audio application, a website, and a web server. Soundjack works in a peer-to-peer fashion well up to seven participants. To support larger groups, mixing servers are needed, which are described in Chapter 3. Second, our Fastmusic Box is introduced, which is a dedicated hardware that helps to identify and reduce significant issues that might hinder a proper NMP experience. Finally, as our experience shows that musicians need to be educated properly in NMP, we report about the users' and our experiences over the last year.

## 2 Soundjack

The Soundjack architecture [10] consists of an audio streaming engine that contains the low-latency audio and video streaming functionality, a website, a web server, and audio mixing servers. Soundjack has been

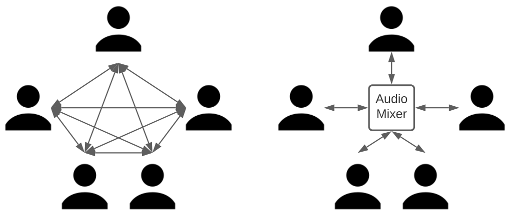


**Fig. 1:** In a mesh peer-to-peer setup (left image), everybody talks with everybody. In an audio mixer scenario, everybody talks only with the central mixing server (right).

available to the public since 2006 and has been used within numerous official network music performances.

As a first step, users have to log into the website https://www.soundjack.eu. The web server handles the connection setup between multiple Soundjack users. It also supports social media features for the end-user, public and private groups, and helps multiple Soundjack users to establish calls between each other.

Because the audio features of a web browser are not fast enough for NMP, we need a dedicated audio streaming engine called the "Soundjack Core" (SJC). It is freely available for OSX, Windows, and Linux. Alternatively, the Fastmusic Box (FMB) can be used instead of a PC-based app. This offers benefits as described in the following chapters. Both the SJC and the FMB achieve a peer-to-peer (P2P) audio transmission latency of 2.5 ms under ideal conditions. This is less than one-twentieth the latency of a web browser or a conventional teleconferencing app.

The SJC supports multiple connections at the same time. The maximum number of P2P connections depends on the available upload bandwidth, the actual system hardware in terms of CPU performance, and the sound card stability. Typically, we do not recommend more than 7 simultaneous calls. Larger ensembles require an audio mixing server as explained below.

## 3 Audio mixing server

When using an audio mixing server, each participant makes a single call to the mixing server. The mixing server receives every single audio stream, mixes them, and sends the mix (minus each individual's own stream) back to each connected participant (Fig. 1).

Compared with the P2P approach, a central mixing introduces additional latency because every stream has to undergo a detour to the server and cannot be sent via a direct link. For example, if ten musicians are in Boston and playing together on a mixing server in Virginia, the audio signals have to travel from Boston to Virginia and back again. The distance between Boston and Richmond, Virginia, is about 880 km (550 miles). On the public Internet, this detour can lead to a total latency of 30 ms or more. In comparison, if these ten musicians were connected via P2P and routed within the Boston area they could experience latencies below 10 ms, well below the threshold needed for making synchronous music. However, if the ten musicians would communicate directly, 45 P2P connections between all musicians need to be established.

It is important to use as centrally located a mixing server as possible with regard to the location of the participants. If located too far away (or even on another continent) the transmission delay would be too high to allow NMP. Thus, Soundjack has numerous mixing servers deployed around the world so that every group, orchestra, or band can select the mixing server at their ideal location. Typically, choosing a mixing server in the center of all participants brings the best results.

The technical requirements of such mixing servers lead to costs of approximately 1€ per hour. Since not all Soundjack mixing servers in the world are needed at any moment of time, we typically halt them and pause their operation. Users can bring them up on demand by a dedicated booking mechanism. To share the resulting costs with the users of the mixing servers we introduced the "Soundjack Coins." One can buy these coins in the Fastmusic webshop and then use them to book a server of choice for a given period.

If an organization permanently needs a mixing server and can buy and operate an appropriate device it is also possible to integrate this server into their individual account. Deploying such a private server will lead to ideal latencies, assuming it is located and managed on the same network segment as the involved peers.

The technical design of a mixing server must take into account that up to 60 incoming audio streams are decoded, mixed, and the mixed signals encoded again. This process must be handled individually for each participant. Since these tasks are computationally demanding and strict real-time deadlines must be met, we consciously spread the involved tasks over the available cores of the server CPU in order to allow parallel signal processing with lowest possible delays. In our implementation, each core handles the network connection, audio encoding, and decoding for 8 participants. However, the audio mixer exclusively reserves one dedicated core. Thus, we ensure that the mixing is always performed under ideal real-time conditions and the mixing result is always punctual. Regarding details and further system specifics, the authors are referring to [10].

## 4 Fastmusic Box

When properly configured, Soundjack enables NMP sessions at excellent quality using an OSX, Windows, or Linux PC with a consumer Internet connection. However, only an experienced user can ensure that the system balances quality and latency. Novices often have to cope with a couple of issues, which hinder a perfect NMP experience.

Based on 20 years of experience in conducting distributed ensemble performances, we understand the most frequent issues that impact the final latency measure. The Fastmusic Box is a software/hardware solution that implements this expert knowledge to make the life of a musician easier. It identifies all common problems, tries to solve them, or at least gives the user hints on what needs to be changed if the issue cannot be automatically addressed. The Fastmusic Box offers a settings page that lists all the issues that might arise. For each issue, a check mark is given. When all check marks are green, a good NMP quality can be guaranteed (see Fig. 2). In this way, the required technical know-how is reduced to a minimum so that more musicians can immediately play with their collaborators at low latency.

The number of potential issues is significant; the FMB checks about fifty of them. They range from networking issues such as latency, loss rate, DNS, NAT, and firewalls, and issues related to the operating system such as drivers, scheduling, and interrupts to the latency of audio interfaces.

The first version of the FMB, which we released on June 1st, 2020, is based on a Raspberry PI 4b and off-the-shelf components. Because of the pressure to deliver a solution for those impacted by the pandemic, we were limited to parts that can be ordered globally

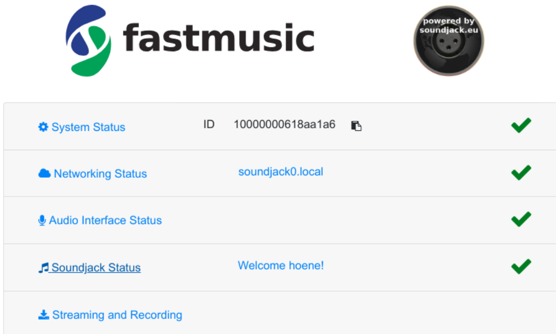


**Fig. 2:** The Fastmusic Box web page with check marks.

**Fig. 3:** The Fastmusic Box 2020 (left) and the Fastmusic Box Pro 2021 with build-in soundcard and display (right).

as illustrated in Figure 3. This ensures that anyone can build an FMB by themselves. One year later, an advanced version of the FMB — the FMB Pro — is available. This includes a display and a built-in audio interface.

There are several advantages of using the FMB over running Soundjack on a personal computer. These include the low price per unit, which allows communities to deploy them at a large scale for the fraction of what computers would cost. Additionally, the FMB allows for value-added custom features such as built-in multi-channel recording and local network streaming. However, the main advantage is that the FMB continuously monitors all networking, system, and operational parameters in a web server app so that any problem can be immediately communicated to the user. If an issue is found, the user is provided with instructions on how to overcome this problem.

In the following, we will go into their details to gain an

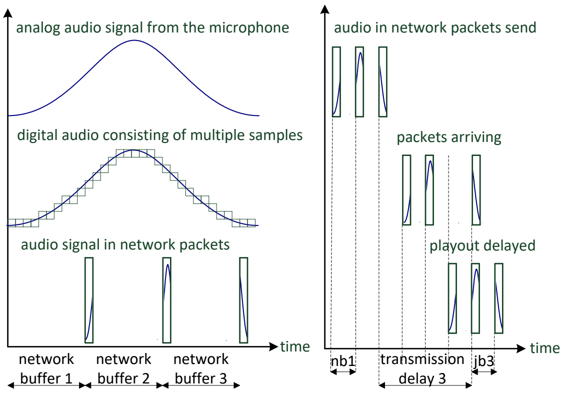


**Fig. 4:** An analog audio signal is digitalized and packed into network packets (right). Audio packets are generated at a regular intervals and transmitted with variable delays. The jitter buffer tries to compensate these variable delays before playout (left).

understanding of what technical requirements need to be fulfilled.

## 4.1 Soundjack settings

The Soundjack app works according to a simple yet effective principle. It records audio sample buffers, compresses them into smaller network packets, and sends them over the Internet to another Soundjack app. There the packets are unpacked, stored for an amount of time in order to accommodate network jitter, and then played back (Fig. 4). Overall, with three parameters this principle can be controlled: the network buffer, the jitter buffer, and the sample buffer as described in the following.

### 4.1.1 Network buffer

The audio signal is digitally sampled with 16 bits at a frequency of 48 kHz. This means that 48,000 times per second a digital value between -32768 and 32767 has to be captured and processed. These values are called audio samples. Classical PC hardware and the Internet itself are not fast enough to process each audio sample independently. Instead, multiple audio samples are put into a packet. Within Soundjack, one can define the number of audio samples per packet: 64, 128, 256, or 512. This parameter is called the "network buffer" in Soundjack terms.

The more audio samples in one network packet, the less bandwidth required. This is because after bit rate, Internet capacity is primarily challenged by the number of packets per second. If a Soundjack call is adversely affected by Internet traffic congestion, increasing the network buffer decreases the likelihood of dropped packets. However, increasing the network buffer comes at the cost of additional latency. One audio sample has a duration of 20.83 $\mu$s. If Soundjack has to wait for 64 samples, this adds a delay of $1.33\ ms = 64/48\ kHz$. 512 audio samples take 10.7 ms, which is already a significant percentage of the 25 ms maximum allowed for a good quality NMP experience.

### 4.1.2 Jitter Buffer

Due to the asynchronous nature of the Internet, the transmission latency of packets from sender to receivers varies. The Internet simply does not provide any guarantees to deliver any IP packet on time. This will always introduce packet to packet variability of transit time, or "jitter". Thus, Soundjack has a jitter buffer, which stores packets on the receiving end for some time. In case the next packet arrives later than the previous one, the playback stream continues unaffected. Otherwise, the late packet would have to be replaced by an extrapolated, suboptimal audio signal.

Again, the jitter buffer comes at costs: Delaying the packets increases the latency. There is generally an optimal jitter buffer size representing the ideal compromise between latency and audio distortion. The Soundjack user can manually set it or allow Soundjack to automatically select an optimal jitter buffer size.

### 4.1.3 Sample Buffer

Finally, the sample buffer size is the number of audio samples the audio interface provides to Soundjack at a time during the input step and the number of samples that Soundjack must provide at a time to the audio interface for playback.

The sample buffer is the base buffer of the network buffer explained above. The network buffer either equals the sample buffer or consists of multiples of it. In any case, the sample buffer determines the audio latency of the base system and therefore should be set as small as possible because then the jitter buffer can be selected more precisely and delays are reduced. Its size is limited by computational resources.

#### 4.1.4 FMB feedback for optimal Soundjack buffer setting

Since the Soundjack settings and the three parameters are not trivial to understand, the FMB offers feedback to select the lowest workable values. If they are not optimal the FMB shows red cross marks. The FMB also provides feedback on a variety of other parameters.

### 4.2 Hardware and Operating System

Beyond the performance of the Soundjack core, we have to ensure that the hardware and operating system are well suited for NMP. Measurements show that Apple's operating systems provide an extremely low audio latency. Windows or Android OS introduce higher latencies by default. Linux is on par with Apple if the Linux system is carefully tuned in terms of real-time kernels and further optimizations. Whereas a modern Apple PC is a good choice for NMP, Android and native Windows PCs without an ASIO [11] audio device are less suitable for NMP because of the audio high latency. Also, the performance of a PC is important. For example, two-core machines such as the MacBook Air of 2015 are not powerful enough to reach low audio latencies.

Concerning hardware, standalone computer modules are available only for Windows and Linux operating systems, either based on Intel x86 or ARM CPUs. For example, Raspberry Pis have been used for NMP in the past [12] but only the latest generation 4 has a low latency Ethernet interface and is computationally powerful enough to be suitable for NMP. Because of the low cost and immediate worldwide availability of the Raspberry Pi 4 we chose it as the platform for the FMB.

It is also important to note that the computational power of the device is well above the minimal needs. For example, if the CPU needs 2 ms to handle a 2 ms segment of the audio signal, the CPU would be just fast enough. However, this would add 2 ms to overall latency as this is the time the CPU needs for the calculations. If, however, the computing power is twice as much as minimally needed, the processing of a 2 ms segment would take only 1 ms. Thus, the computational delay would be reduced by 1 ms. In this way, additional computational power further reduces latency.

Multiple operating systems are available for Raspberry Pis. For example, real-time operating systems (RTOS) such as FreeRTOS or ChibiOS/RT have the potential to allow for precise audio scheduling in order to further minimize audio processing latency. This scheduling is crucial because otherwise audio dropouts in resulting clicks would be added to the audio playout.

Those RTOS typically don't support additional software components such as a web browser needed for Soundjack. Thus, we decided to use the Linux distribution Raspberry Pi OS extended by a real-time Linux kernel and customized scheduling. However, during the pandemic, it has been challenging to ensure that the audio scheduling of the Linux operating system is always on time. This has been the greatest challenge in the entire FMB development. We assigned each process in the system a specific latency, assigned audio processing parts to specific CPU cores, and improved the interrupt handling to overcome these issues.

### 4.3 Updates

We have improved and updated the software constantly over time. New features have been introduced, bugs fixed, and security issues solved. To manage both functionality and security, the FMB updates itself automatically every night (and after every start) to get the latest version from the operating system provider. For other system settings, we use the software configuration management (SCM) tool Puppet [13] which ensures that all settings, files, and apps are always up-to-date. Puppet has been used for 15 years to configure servers but has also proven reliable to update the FMBs. This system ensures that all FMBs are always up to date.

These automatic updates have allowed us to keep the development pace at a high level. Especially at the beginning of the 2020-21 academic year, we provided new features to the FMBs on a weekly basis. Whereas a similar update rate is known for websites, updating embedded devices so regularly is rather unusual. However, it was crucial to provide support to users as quickly as possible.

### 4.4 Audio Interface

Because of time constraints, we were not able to include an integrated sound card into the first version of the Fastmusic Box released in June 2020. Only the FMB Pro, released a year later, includes a customized sound card with phantom power, XLR jacks, and headphone connectors. Thus, we had to initially support

USB audio interfaces. Fortunately, numerous musicians and teachers already own a USB audio interface.

While almost all USB audio interfaces work with the FMB, they differ significantly in terms of latency. Depending on the model and brand it takes a different amount of time for the digital audio to be transmitted via the USB bus, converted to analog, digitized, processed, and transmitted to the PC again. Some audio interfaces (and USB microphones) need more than 25 ms for these steps thus they are not suitable for NMP.

As we cannot control which audio interface is used with the FMB, we carefully monitor the type and brand of connected audio interfaces. For this, we used the USB ID which is unique for each manufacturer and model. If the model's latency is unknown, the user is requested to conduct a so-called latency test. This is accomplished by connecting the input and output of the audio interface via an analog cable and setting the gain so that the amplification of input and output is at approximately 0 dB. After pressing a latency test button on the GUI of the FMB it sends out a Dirac delta impulse (a clicking sound) and waits until it receives this impulse via the audio interface input. The delay between sending and receiving is the audio latency of the audio interface. If possible we send a block of 64 audio samples via the USB bus. Because of this, a packetization delay of $64/48000\ Hz = 1.33\ ms$ is added to the latency measurements.

If the audio round trip time is less than 10 ms, we assume that is fast enough and the audio interface gets a green check-mark. Otherwise, it is not suitable. Any audio interface that has been tested before by somebody else is stored in FMB and reported to all users. In that case, a latency test is not required.

During the last year, we collected the latency data of more than 55 sound cards and audio interfaces [14]. The fastest integrated sound card at 64 samples is a Hifiberry ADC+DAC pro with 3.1 ms and the fastest external audio interfaces at 64 samples are the Scarlett 2i2 and the RME Babyface Pro (both 5.8 ms).

Aside from latency, some consumer audio interfaces presented challenges. For example, the RPi 3.0A power supply was underpowered, the USB cable shipped with an audio interface was of poor quality, or the audio interfaces were beyond the required technical specification of USB. Because of these potential issues with USB audio interfaces, we see it as beneficial to have a low-latency integrated sound card. Not only is the latency lower, but by controlling the hardware, potential compatibility issues are removed. The FMB Pro is equipped with a Hifiberry ADC+DAC Pro sound card that addresses these issues. As the sound cards for the Raspberry PI hardware do not support the automatic loading of the correct drivers, sound card detection was added to the FMB.

Also, for all USB audio interfaces, the default settings of the audio mixers have been good so that no issues have been reported. Only for the integrated sound card, specific driver settings are required to select the right in- and output ports to reduce the filtering delays of ADC and DAC and to switch on and off the microphone's bias voltage. In order to relieve the user from these problems default audio mixer settings can be applied.

### 4.5 LAN network

The local network quality is critical for a good NMP experience. Wifi or cellular networks typically sometimes show a high transmission latency because the quality of the wireless physical transmission path can not be guaranteed, at times greatly increasing network transmission jitter. Also, the air is shared amount multiple electromagnetic senders, which may interfere with each other. Thus, it is not advised to use a wireless transmission medium for NMP. Instead, a wired Ethernet cable connection is recommended. For this reason, we disabled the FMB WiFi adapter for Internet access and require using Ethernet. In the FMB, a Wifi access point is only provided to remote control the FMB or access the Soundjack website.

In addition, sometimes home routers introduce latency. To identify this problem, we search for the first IP address in the transmission path towards the Internet and use an ICMP ping message to check its latency. If it is above a few milliseconds, we report this to the user with a red check mark so that they can optimize their local network.

### 4.6 Network address translation and Firewalls

Numerous access routers implement a network address translation (NAT), which maps a local (LAN) Internet Protocol (IP) address of the devices to the globally unique IP address. The Soundjack app needs to communicate to other Soundjack apps, even if both are behind such a NAT. However, in specific setups, certain NATs do not allow the forwarding of an incoming UDP stream. Therefore, the FMB verifies the behavior of a NAT to indicate whether the NAT has to be replaced or reconfigured. Fortunately, in more than 95% of cases, we have experienced a positive NAT behavior. Besides a NAT, Firewalls frequently prevent the FMBs from communicating with each other. Such a scenario is also detected and marked with a red cross mark.

### 4.7 Access Network

Once the LAN is passed, the next stage is the access network. Access networks differ significantly regarding technology, bandwidth, and latency. DSL, cable, fiber, or wireless transmission technology are available. The maximal available bandwidth is typically communicated in Internet provider contracts and advertisements. However, for one single mono audio stream Soundjack requires less than 1 MBit/s which nowadays is available by default. However, it also needs a low latency IP packet transmission in the access network. Internet providers typically don't communicate this parameter so that one can only guess if a not yet installed Internet access is fast enough.

Global Internet companies such as Google, Cloudflare, and Quad9 have installed Domain Name Service (DNS) servers virtually all around the globe because low DNS lookup times are important for a smooth browsing experience. As we have experienced that these DNS servers are among the servers in the Internet which respond fastest, we use them to measure the latency of the access network: The FMB pings with an ICMP message the DNS servers 1.1.1.1, 8.8.8.8, and 9.9.9.9 to figure out how much time is spent in the access network. If these pings round trip time are above 15 ms we mark the Internet access network as yellow, 25 ms or more is marked red.

If the Internet access network is too slow then the NMP user has to change their Internet provider. For example, some universities do have two Internet access links: one for bandwidth-intensive applications, the other for local, low latency connectivity. In such cases, we have suggested the application of a service-specific routing so that NMP related traffic goes via the low latency link and other Internet services such as broadcasting via the high bandwidth link.

All of the above ping measures are graphed in real-time on the FMB settings page to aid in troubleshooting.

### 4.8 Global routing

The Internet backbones typically consist of fiber optic links. Light travels in fiber optics cables not at the maximal speed of light ($c = 3 * 10^8\ m/s$), but at about two-thirds of it. Thus, within 25 ms about 5,112 km is traveled under ideal conditions. However, conditions are rarely ideal as routers have to be passed, routing does not follow the shortest path, and other effects according to [15].

Again, the FMB and Soundjack can only point out the problem that the latency or the geographical distance between two NMP participants is too high. As we continue to work to reduce other sources of latency as far as possible, larger distances will be tolerated.

## 5 Lecturing

Much of the challenge of the past year has centered on educating musicians in the use of NMP. Historically, NMP has only been used by a niche of technically interested and tech-savvy people — not at all by the masses of musicians and music teachers. If the ideal solution requires both usable technology and technology-competent users, implementation of NMP at the start of the pandemic was equally limited by the users themselves. Most musicians were not aware of, let alone trained to optimize NMP technology.

Even the NMP technologies that this population discovered had not been optimized for widespread deployment. Just in 2019 two of the authors completed the German research project fast-music which provided funding for the fundamental basics of the FMB technology [10]. The results could be considered research prototypes but not products, which require a significant amount of polishing and optimization to bring to market. Other providers of NMP technologies [6] such as Jacktrip or Jamulus were in similar positions at the beginning of the pandemic. Commercial players such as JamKazam were not investing because the pre-pandemic market was too small to be profitable.

Thus, we experienced difficulty bridging the gap between NMP researchers on one side and musicians on the other. The authors of this publication came to know each other only last year, whilst Ian Howell was looking for a suitable NMP solution he could use in a variety of settings at the New England Conservatory of Music (NEC) in Boston. At NEC, Ian Howell's technical team introduced NMP to the campus, selected a proper NMP technology, wrote manuals and advanced workflows [16, 17], created audio/visual media demonstrating its utility, identified bugs, and issues with the FMB and Soundjack, provided support via Facebook groups and direct email lists, and conducted multiple webinars to teach fellow musicians.

Over the summer of 2020, we saw that musicians, especially those in the NEC College Voice Department and our Preparatory Jazz Department, were willing to engage with this new technology. This led to deployments of more than 100 FMBs and seventeen PC-based on-campus room installations leading up to the fall. This allowed our community members to continue rehearsing, teaching and coaching unhindered by logistical restrictions. We were able to teach again, to learn again, or to simply enjoy playing music together as though the pandemic were not a limiting factor. We incorporated Soundjack into live performances, performance classes, and recording sessions on a regular basis. Even our most technology-averse faculty members acknowledge how useful NMP is to the daily work of music education.

As a consequence of our work and public advocacy, many universities in the US were able to use Soundjack and the FMB during the 2020-2021 academic year. At NEC alone, dozens of Thousands of musicians have been using the free-of-charge Soundjack platform to play music together. During the last months of social distancing, hundreds of music students and teachers have used the FMB on a daily basis to continue singing, playing, and rehearsing music together.

Furthermore, we have also attracted interest from advanced choirs such as the famous Gewandhaus choir, Leipzig, to apply our FMBs in combination with our server-based streaming solution to allow up to 25 singers located at their homes to perform together at the same time as if being located in the same room. In this context, we used our audio mixing server in Frankfurt a.M., which serves as the peering point for most German Internet providers anyways, and in that regard, a minimum of additional latency was to be expected. As a consequence, we indeed saw an average latency of 15 to 20 ms for each participant to the server which allowed us to perform medium-tempo pieces.

These mixing server-based solutions offer additional advantages for even smaller ensembles, including the simplification or elimination of home router configurations and the decreased CPU load per peer computer. Anyone who remains interested in using NMP in either a transitional or post-pandemic world may find the mixing server option to be as low cost and as technically simple a solution as is available.

Ian Howell and his team at NEC were awarded the American Academy of Teachers of Singing (AATS) Award for COVID-19 Response. AATS states that, "Howell's work with low-latency platforms and associated technology, and broad dissemination of instruction in its use, has permitted widespread applications that have allowed us to teach and perform successfully in remote settings. These innovations will have a lasting impact on our profession" [18]. What this lasting impact is will unfold over the coming years. However, based on the experience of the authors, NMP appears more likely to find its way into music technology curricula and into the lives of professional, pre-professional, and avocational musicians than disappear.

## 6 Conclusions

With rising vaccination and sinking infection rates, the need for NMP as a complete replacement for in person music-making is thankfully diminishing. We can clearly see the number of users declining already. However, as restrictions lift and life moves toward a new normal, NMP will become a tool of convenience and collaboration: Instead of driving miles to meet fellow musicians, one can simply call them via the Internet just as in a teleconferencing call. Within the extant music education and performance industries, numerous, daily use cases would benefit from the added convenience and cost-saving. E.g. countless voice students travel between East coast American cities each day only to take a lesson in their teacher's living room anyway. They may as well take the lesson from their own living room, save the travel costs, and take more lessons per month with the money saved. Furthermore, NMP will continue to mature in terms of technology and usability.

On the other side, we have seen that there is a real need to teach music students the basics of audio technology, including NMP. Advanced, low-latency solutions are useless if a participant does not understand how a microphone preamp works. We can assume that in the coming decades, most musicians will at some point be exposed to NMP. Introducing this early in a music student's education will require curricular changes, but the authors argue that it is as important to understand NMP as it is to understand any other music-adjacent technology. Additionally, we would like to see that Internet providers advertise the latency of their Internet access as prominent as the bit rate - and optimize their offerings accordingly.

Nevertheless, regardless of all the benefits NMP has brought us over the last year, we are more than happy that these days we can start meeting our fellow musician friends in person again.

## 7 Acknowledgments

We are very thankful to the Soundjack users and community for providing kind, helpful, and constructive feedback over the last years. We also appreciate their patience as we have worked toward the stability and usability of the Soundjack and FMB NMP solution. In particular, we have to mention Bob Avant, Eden Casteel, Kayla Gautereaux, Urs Gehrig, Sarah Hoene, Bernd Keul, Matthew Chellis, Dana Long Zenobi, Christoph Kuhr, David Liao, Nicholas Perna, Bonnie Sneed, and Oliver Worthington for their constant contributions during the last year. And to our families and children who were the most patient of all.

## References

[1] Vance, D., Shah, P., and Sataloff, R. T., “COVID-19: Impact on the Musician and Returning to Singing; A Literature Review,” *Journal of Voice*, 2021, ISSN 0892-1997.

[2] Miller, S. L. et al., “Transmission of SARS-CoV-2 by inhalation of respiratory aerosol in the Skagit Valley Chorale superspreading event,” *Indoor Air*, 31(2), 2021.

[3] Naunheim, M. et al., “Safer Singing During the SARS-CoV-2 Pandemic: What We Know and What We Don’t,” *Journal of Voice*, 2020.

[4] Echternach, M. et al., “Impulse Dispersion of Aerosols during Singing and Speaking: A Potential COVID-19 Transmission Pathway,” *American Journal of Respiratory and Critical Care Medicine*, 202(11), 2020.

[5] Howell, I. et al., “Preliminary Report: Comparing the Audio Quality of Classical Music Lessons Over Zoom, Microsoft Teams, VoiceLessonsApp, and Apple FaceTime.” https://www.ianhowellcountertenor.com/preliminary-report-testing-video-conferencing-platforms, 2020.

[6] Rottondi, C. et al., “An Overview on Networked Music Performance Technologies,” *IEEE Access*, 4, 2016.

[7] Carôt, A. et al., “Creation of a Hyper-Realistic Remote Music Session with Professional Musicians and Public Audiences Using 5G Commodity Hardware,” *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 2020.

[8] Carôt, A., Renaud, A., and Verbrugghe, B., “Network Music Performance (NMP) with Soundjack,” in *Proceedings of the 6th NIME Conference*, Paris, France, 2006.

[9] https://www.musicforall.org.uk/, 2021.

[10] Carôt, A. et al., “Results of The Fast-Music Project – Five Contributions to The Domain of Distributed Music,” *IEEE Access*, 2020.

[11] Ye, D. et al., “Measurement and analysis on audio latency for multiple operating systems,” in *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, 2018.

[12] Meier, F., Fink, M., and Zölzer, U., “The jamberry-a stand-alone device for networked music performance based on the raspberry pi,” in *Linux Audio Conference 2014*, 2014.

[13] https://puppet.com/open-source/, 2021.

[14] https://symonics.com/tested-soundcards, 2021.

[15] Briscoe, B. et al., “Reducing Internet Latency: A Survey of Techniques and Their Merits,” *IEEE Communications Surveys & Tutorials*, 18, 2014.

[16] Howell, I., “SoundJack Guide,” https://www.ianhowellcountertenor.com/soundjack-comprehensive-user-guide, 2020.

[17] Howell, I. and Kelly, M., “A Real Time, Distributed Performance Class Using SoundJack and Zoom,” https://www.ianhowellcountertenor.com/distributed-performance-class, 2020.

[18] https://www.americanacademyofteachersofsinging.org/, 2021.