Distributed Network Music Workshop with Soundjack

Tonmeistertagung 2008 / Leipzig, Germany

Alexander Carôt[1]
Christian Werner[1]

[1] Institute of Telematics, University Lübeck, Germany

Abstract

Playing live with someone abroad can be considered as a major challenge for musicians and sound engineers likewise. Due to cognitive, technical and purely musical problems and restrictions it had so far been impossible to reproduce a realistic rehearsing scenario as if in the same room. However, since nowadays the Internet offers sufficiant bandwidth and reliability it can satisfy the high demands and actually fulfill the extreme time critical restrictions. In order to achieve the most decent latency and quality of a network, the Soundjack software has been developed and applied in terms of a musical telepresence. In this workshop the authors will demonstrate what problems sound engineers and musicians have to overcome in context with audio- and network delays and how alternative high delay solutions can be applied. As a practical showcase three musicians in Lübeck/Germany, Leipzig/Germany and possibly further locations will be connected via the Soundjack software in order to play music under different delay conditions. In that context relevant Soundjack features will be explained and applied to the current network condition.

1 Introduction

Until the late 1960s wide area networks were commonly available in terms of voice telecommunication only [14]. Asynchronous data networks played no significant role in that context as they were mainly intended for data retrieval in local area networks of companies. Hence, in the past we could clearly distinguish between synchronous telecommunication networks and asynchronous industrial data networks. Although in the first decade of the 21st century this separation still exists, it is fairly not as strict as it used to be: Following McLuhan in [17], that sooner or later every kind of medium will finally end up in a single multi purpose medium, customers more and more asked for multimedia and communication services for audio and video on one hand and at the same for services in data retrieval and non-realtime communication on the other hand. In order to be prepared for the new mixture of real time and data services telephone providers consequently had high hopes in a new broadband ISDN (B-ISDN) [21], which preferably applied the isochronous ATM [11] in its data link layer. Things, however, developed differently than expected as people became more and more attracted to the Internet, which nowadays offers an enormous amount of data hosted on globally interconnected machines and can already be considered as the major commonly available source of information. Since the Internet has such an impact on our daily lives, the idea of a network that covers any desired service, has become attractive to users and the industry as well. Hence, and despite the drawbacks of asynchrony, in 1994 first services offering Internet radio came to life, followed by Internet telephony (VoIP) and videoconferencing around 1998 [23]. These services have been improved constantly and about 10 years later they have become extremely reliable and a generally accepted form of realtime communication [15]. The famous VoIP tool “Skype” [3] represents a good example for this tendency. The common approach to compensate the effect of network jitter is the application of a jitter buffer at a receiver‘s end: By storing up an amount of audio packets in the network queue the audio process can still provide a solid playback in case of late arriving packets. The drawback of this principle is a higher latency as packets are not processed right after the reception [16].

In terms of music the Internet has so far mainly been used for the interchange of audio tracks in the music production process. The realistic music interplay, however, represents a more complex field of application: Cognginitive, technical and purely musical aspects make high demands on a network music performance that should fulfill conditions of a realistic rehearsing scenarios in the same room. In turn the idea of a network music performance has generally been considered as an impracticable application as the total one-way delay between two performers is supposed to remain below 25 ms [4]. Figure 1 shows the total path a signal has to pass in order to be transmitted via the Internet. Each stage is marked in either white, light grey or dark grey color. White indicates no additional delay, light and dark grey indicate slight and significant delay respectively. On top of the natural path, the electronic path and the digital path, the Internet might add significant delays due to detours introduced by the routing of packages and due to jitter buffers in order to compensate delay variations. Based on these facts we considered to develop a tool for the Internet, which suits this transmission structure and allows to verify a system‘s soundcard setting and the actual network parameters in terms of an individual parameter optimization.

2 Distributed music with Soundjack

Soundjack was developed by Alexander Carôt at the University of Lübeck/Germany and is a low delay audio streaming tool for the Internet, which grabs the soundcard‘s audio chunks and transmits them via the UDP/IP protocols [19] [18] as directly as possible to a remote destination. The transmission and delay conditions only depend on the applied hardware devices and the actual network structure Soundjack is used with. In order to precisely figure the latencies for a stable network audio stream, it allows the adjustment of any relevant audio and network parameter, and in that context it is divided into a network section, a traffic generator and an audio section, which will be explained in the following subsections.

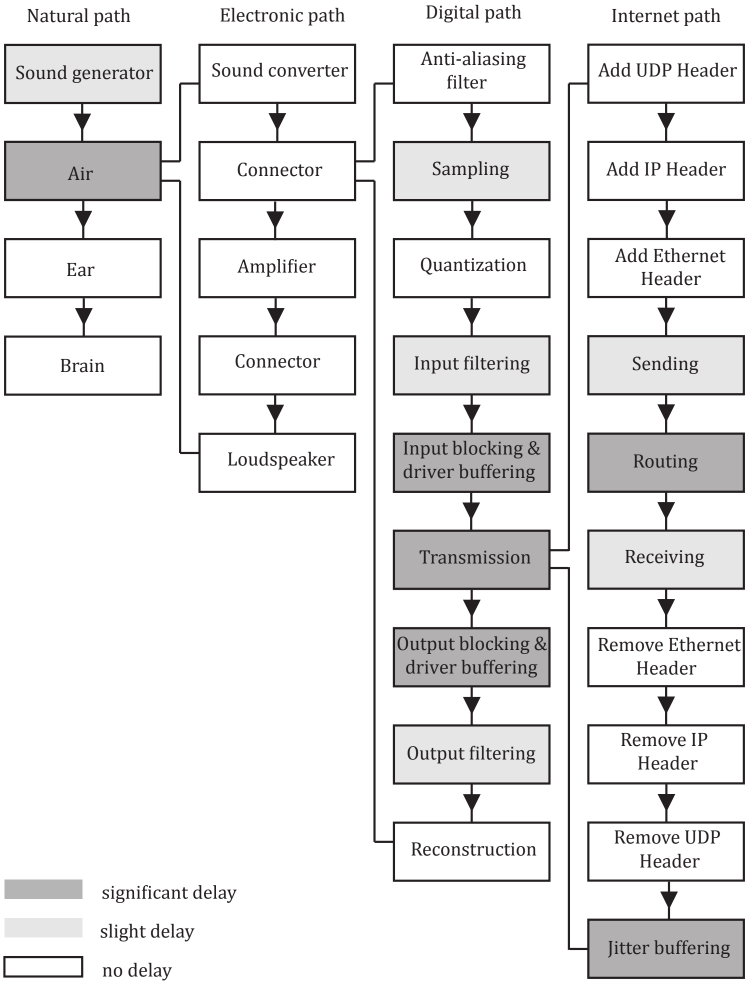
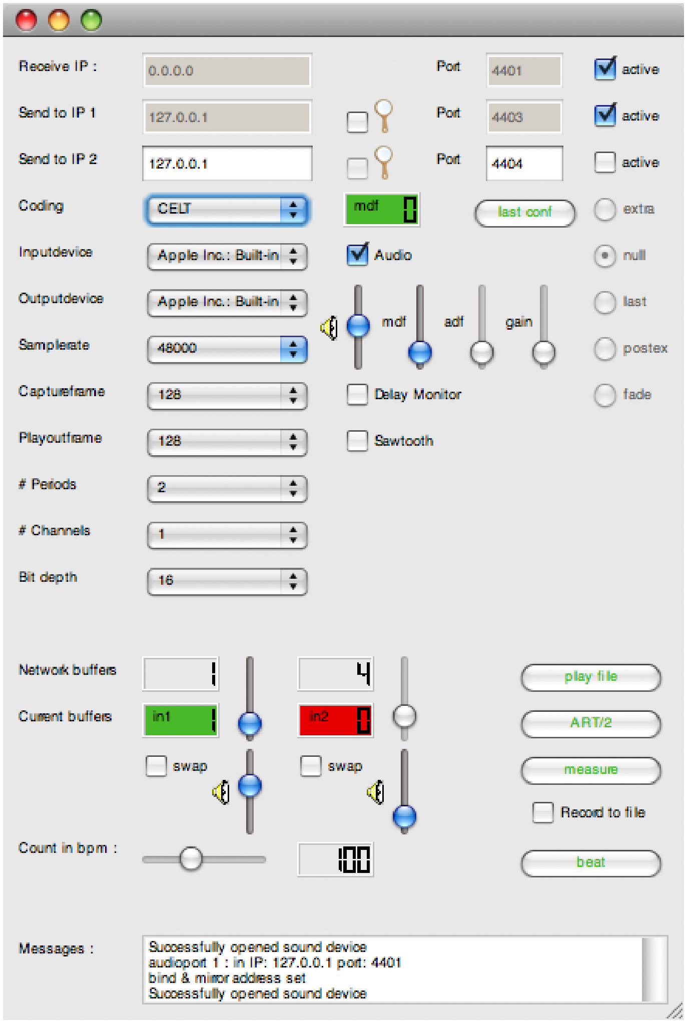


Figure 1: Total signal path

2.1 Network Section

The network section is controlled by the user using the first three lines of the GUI shown in figure 2, where IP addresses and ports have to be specified.

In order to be able to send and receive data an endpoint UDP socket has to be bound to the IP address of the current network interface. In ideal case it is bound to an external IP address in the public Internet. If the endpoint is located behind a network address translator (NAT) [12] it will have to bind to a private IP address. In this case the next router‘s port forwarding has to be enabled: With port forwarding remote peers can use the router‘s external IP address as their destination which will forward all data on specific ports to the internal endpoint. This configuration generally happens via a web interface, in which the respective ports have to be specified. Hence I declared ports for the respective sockets. By default a socket bound to port 4401 sends



Receive IP : 0.0.0.0 Port 4401 ☑ active

Send to IP 1 127.0.0.1 □ Port 4403 ☑ active

Send to IP 2 127.0.0.1 □ Port 4404 □ active

Coding CELT

Inputdevice Apple Inc.: Built-in

Outputdevice Apple Inc.: Built-in

Samplerate 48000

Captureframe 128

Playoutframe 128

Periods 2

Channels 1

Bit depth 16

☑ Audio

□ Delay Monitor

□ Sawtooth

○ extra
● null
○ last
○ postex
○ fade

Network buffers 1 4

Current buffers in1 1 in2 0

□ swap □ swap

□ Record to file

Count in bpm : 100

Messages :

```
Successfully opened sound device
audioport 1 : in IP: 127.0.0.1 port: 4401
bind & mirror address set
Successfully opened sound device
```

Figure 2: Soundjack main interface

data to the IP address and port specified as the first remote destination, while another socket bound to port 4402 sends data to the IP and port specified as the second remote destination. A further socket bound to port 4403 receives data of a first remote player, while socket port 4404 receives data of a second remote player. In terms of buffering incoming packets a FIFO-buffer for each input channel has to be applied, which stores up a desired number of packages before the application actually reads from this buffer. Furthermore another socket bound to port UDP port 4405 is the so called “mirror” which reflects all received data back to its sender. This is useful for testing the actual network conditions in a situation, where the remote partner is not available. Each IP address and port combination can be enabled or disabled with the respective checkboxes at the line‘s end.

As a special feature Soundjack provides audio transmission via ICMP packets by enabling the mirror icon on the right hand side of the IP address field. Once it is checked the ICMP echo service can be used. This has the same effect as if the UDP mirror port would be used, except that the ICMP echo request does not require a running instance of Soundjack on the remote peer. Furthermore this principle allows to use any hop on the route to the destination as a reflector in order to figure problematic network segments [8].

2.2 Traffic Generator

Before establishing an audio connection between two peers, it is useful to verify a network link in terms of throughput, delay and jitter. In that context we also developed a traffic generator, for which further ports are declared: Once the “traffic generator” checkbox in the network section is checked and “dest1” in the traffic generator is enabled, Port 4406 sends one UDP packet to IP1 and Port1 specified in the network section. Having the feedback checkbox enabled, the sender uses the remote side‘s mirrorport as its destination which reflects all data back to its sender. In this case the one-way delay time is shown graphically in the delay gauger. Port 4407 works respectively if the second destination is enabled. In the default setting one packet of 10 bytes is sent each 100 ms. The amount of bytes per packet and the sending frequency can be adjusted manually while the actual amount of data in kBit/s is displayed below. By observing the delay display the user will already get an impression, in how far the current link is able to transmit the data stream. However, besides this we applied further jitter displays, which explicitly provide information about the remote side‘s jitter and the local jitter. The less variation is displayed, the more stable the network link is in terms of better audio quality.

2.3 Audio section

The most important settings are located in the soundcard section since this is where the main decisions in terms of bandwidth, overhead and latency take place [5]. Firstly, one has to choose an available soundcard for the audio processing. If desired the transmission and the reception can be allocated to two separate cards. This is followed by the choice of the sample rate and the audio framesize. The higher the samplerate and the lower the sample framesize the smaller is the delay – as illustrated in figure 3. Furthermore a reduction of the delay implies a larger amount of packets per second. Each packet includes the so-called packet overhead, which consists of at least the UDP and the IP header, and the more packets sent, the larger this overhead becomes. Depending on the used network technology other protocol headers have to be added. Especially in DSL networks this can add a significant amount of overhead bandwidth as described in [5]. In case the available bandwidth capacity does not suffice it is possible to decimate audio blocks by factor 2, 4 and 8 in the “Coder” field. In terms of a better signal quality alternatively the Fraunhofer ULD [5] or the open source CELT [22] codec can be applied.

With the “number of periods” - field the number of internal soundcard buffers can be adjusted. A number of two is the lowest value soundcards can work with. In case of unstable behavior with corresponding audio dropouts the value has to be increased

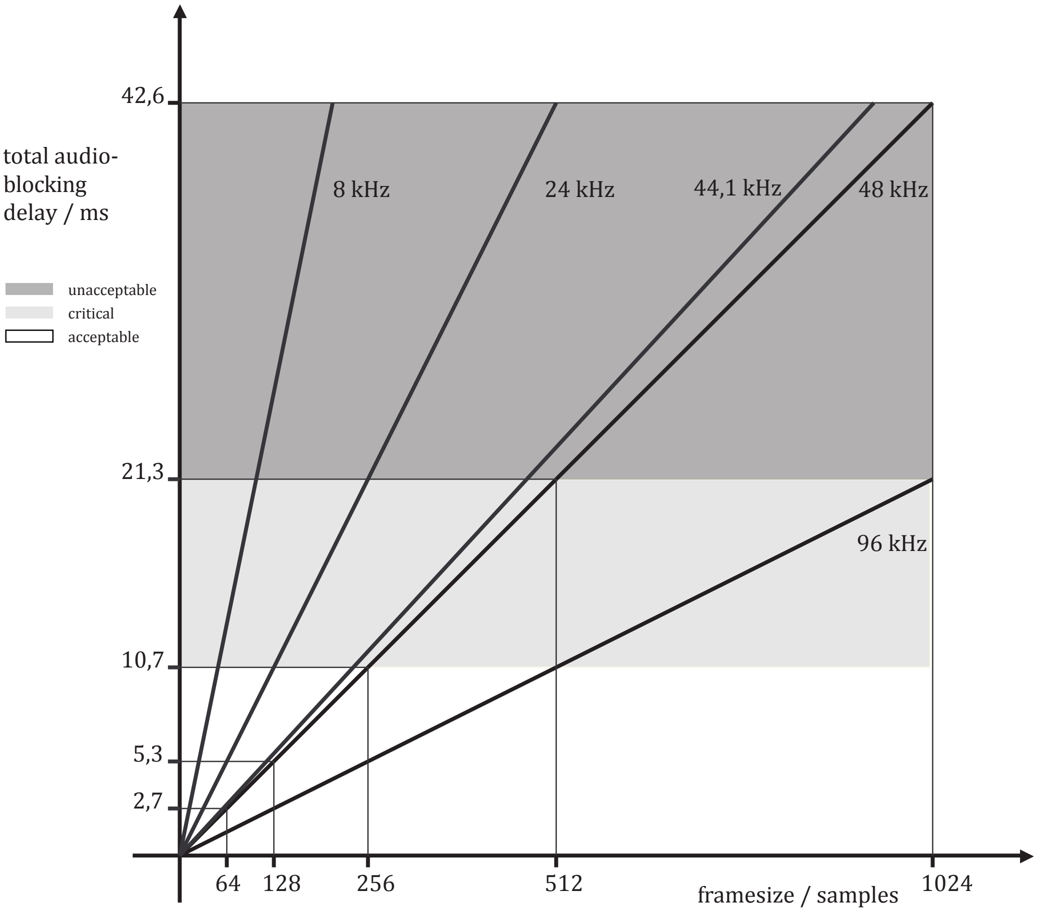


Figure 3: Total audio blocking latencies

which also results in additional latency. Regarding the loopback mode it is possible with Soundjack to adjust the locally produced sound with the “you” volume fader. Furthermore this local playback can be artificially delayed with the “mdf”-fader (manual delayed feedback) by an adjustable number of periods. The “adf”-fader (automatic delayed feedback) provides an automatic adjustment with the roundtrip delay. Both faders can be useful if the delay is too large for playing in an uncompromised, realistic manner (c.f. section 3). Concerning a remote partner‘s signal the mixer section below the soundcard settings allows the adjustment of the input level of each participant via the “In1” and “In2” input level faders. Above these faders one can adjust the jitter buffer for the according channel with the network buffers fader.

Rather than the pure and unbuffered network delay the real audio delay between two endpoints would be of interest, which includes latencies introduced by the buffering and soundcard playback. In that context we declared the term “ART” as an abbreviation of the audio roundtrip trip time. This value can theoretically be retrieved by starting a time measurement when marking and sending one specific sound buffer packet to the remote destination. After passing the remote end‘s signal chain the soundcard notices the input of this specific sound buffer packet and sends it back to the sender. At the moment the marked audio buffer has reentered the origin‘s soundcard, the time measurement is stopped and in turn the roundtrip time can be retrieved. As a rough estimation the one-way audio delay is represented by a division of the roundtrip time by 2. If the user likes to record the processed audio stream the “rec” checkbox writes the mixed audio streams into an audio file called “mix.raw” in raw format. In case the remote partner uses a Power PC Macintosh Comptuter (PPC) while the other machine is a regular PC or an Intel Macintosh PC, the swap button has to be enabled since both machines send bytes in a different order. The swap button reorders them and thus makes the buffer readable. Leaving the swap button switched off will lead to heavy distortions on the receiving end. Furthermore Soundjack provides a metronome users can apply musically or as reference test signal. Finally the “last conf“ button restores the previous session‘s parameters in order to allow quicker configuration.

3 Interaction categories

There is no doubt that the Internet as an asynchronous medium is not the predestinated choice for exchanging real-time audio data and especially not for low delayed audiostreams with their according low block sizes. Nevertheless with constantly increasing bandwidths and respectively increasing transmission speeds the effect of network jitter has been loosing significance over the years. Furthermore due to the Internet‘s large distribution and its strong sociocultural establishment the Internet already beats the outdated synchronous telephone network so that further investigations in distributed music systems will use the Internet as the prime transmission system of choice. However, apart from the applied network technology it is clear that even in case of a direct fiber optic link between two players the maximal distance might not exceed 5250 km in order to let the pure network delay undershoot the 25 ms delay threshold. Nevertheless on top of that cable detours due to signal routing, device processing delays and soundcard delays decrease this maximal playable distance typically below 1000 km. Especially in a worldwide scenario delays of more than 100 ms are likely to expect and thus making distributed music impossible. However, taking further cognitive aspects of musical interaction into account it is possible to consider compromised ways to cope with latencies beyond the 25 ms threshold.

In that context we generally have to distinguish between a solo instrument and a rhythm instrument. Apart from that the placement of the so-called "rhythm section" is of significant importance. As an example with drums, bass and saxophone, a simple case is present in which drums and bass form the "rhythm section" while the saxophone player represents the "solo section". Though of course the solo section and any musician must have a sense of rhythm, it is basically the interplay of bass and drums, which forms the essential fundamental groove of an ensemble and which allows other solo instruments to play upon. In this scenario the saxophone player relies on and plays on the groove that is produced by the rhythm section [7]. Due to the fact that rhythm and synchrony are the main fundament of groove based music, the following subsections put emphasis on rhythm based instruments and the groove building process. In classical music things are more complex: Here we usually cannot precisely distinguish between "rhythm" and "solo" sections [7]. Anyhow, in most pieces of classical music an analogical categorization is feasible, but should be more fine-grained and more dynamic. Also the concept of a conductor has to be considered here. In the following – in order to present our concepts as clear as possible to the reader – we will focus on applications in the field of rhythmical music and continuously use the according terms "solo section" and "rhythm section". Based on the actual delay between two player, we can divide the possibilities of a musical interplay into four main categories.

3.1 Realistic Interaction Approach (RIA)

A realistic musical interaction, as if in the same room, assumes a stable one-way latency of less than 25 ms [4] between two rhythm-based instruments such as drums and bass. In this scenario both instruments‘ grooves merge into each other and the real musical interplay can happen [9]. From the perceptual point of view the delay appears to be as not existing, which is similar to musicians playing with a maximal physical distance of about eight meters in a rehearsing space, where the speed of sound is the limiting time delay factor. The RIA is the only approach professional musicians accept without any compromise since it is the only scenario, which exactly represents the conventional process of creating music in groups or bands. Beyond this threshold of 25 ms, the groove-building-process cannot be realized by musicians anymore and thus different compromises have to be accepted [7]. Figure 4 shows that below 25 ms delay both players are able to play at the same instant and receive each other‘s signals as if no delay was existent.

Due to technical difficulties in applying the required RIA conditions, RIA has so far not turned into a commercial entity but has mainly been examined in research projects, such as SoundWIRE by Chris Chafe of CCRMA [10] and our Soundjack system [6].

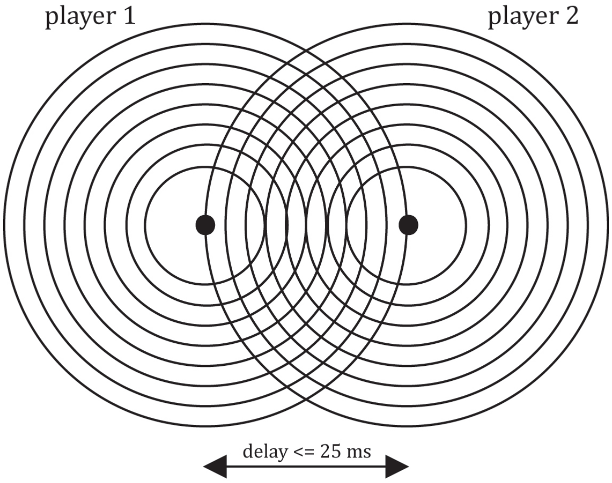


Figure 4: Realistic interaction approach

3.2 Master Slave Approach (MSA)

Assuming an attendance to compromise and to step back from musical perfection and ideals, it really is feasible to perform with two rhythm-based instruments such as drums and bass, even when exceeding the 25 ms threshold – simply if one of the musicians keeps track of his rhythm and does not listen to the incoming high delayed signal anymore. In that situation the remote side can perfectly play to the incoming signal since the other side doesn't care about the response anymore – a change in the musical interaction is happening, which here is called the "Master-Slave"-Approach. The first musician takes the master role since he is producing the basic groove while the remote musician simply relies on it and hence takes the slave role [20]. Of course the higher the delay, the more difficult the ignorance of the delayed input can be realized by the master since shorter delays will easier establish a musical connection to the previously played notes. In terms of delay MSA generates no latency and perfect sync on the slave's side but on the other hand it delays the slave with the roundtrip delay on the master's side. While the slave musically depends on the master but has a perfect sync, the master has musical independency but an unsatisfying sync [7]. Figure 5 shows a situation with a delay beyond 25 ms delay between two players. Due to the high delay the slave has to wait playing until the master‘s signal has arrived, which finally leads to a roundtrip delay on the master‘s end.

In general the master role is taken by a rhythmic instrument in order to let solo instruments play on its groove in slave mode. An exception can happen when a rhythmic instrument suddenly starts with a solo part. In this case it will require the other instrument to take over the leading rhythmic role, which in turn leads to a switch of roles. MSA can be applied with any system that allows the transmission of realtime data. This could be a tool for IP telephony or videoconferencing, which does not put emphasis on low delay signal transmission, but as well high speed audio transmitters

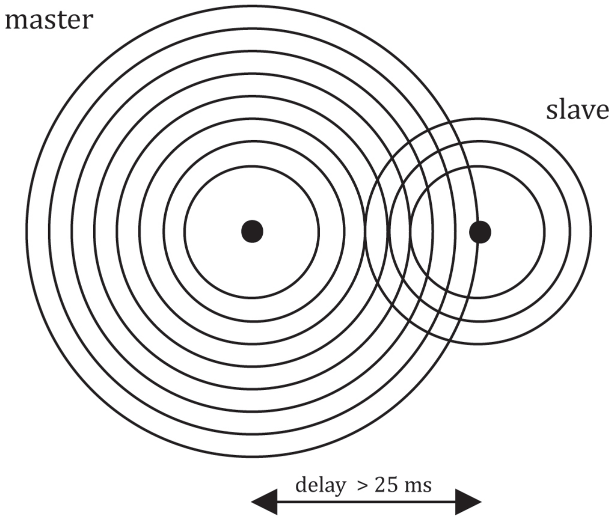


Figure 5: Master/slave approach

in an intercontinental setup. In the latter case the main source of latency is the long physical distance.

3.3 Laid Back Approach (LBA)

The Laid-Back-Approach is based on the "laid back" playing manner, which is a common and accepted solo style in jazz music. Playing "laid back" means to play slightly behind the groove, which musicians often try to achieve consciously in order to make their solo appear more interesting and free. The Laid-Back-Approach is similar to the Master-Slave-Approach and is mainly determined by the number of participating instruments and their role. As previously mentioned, two rhythm-based instruments separated by delays beyond 25 ms have to play with MSA but in case of one of the instruments being a solo instrument, the situation changes. Exchanging the drums with a saxophone in the example scenario results in a remote rhythm/solo – constellation in which the bass represents the rhythm instrument and the saxophone the solo instrument. Since the bass now has no rhythmic counterpart anymore, it alone takes the responsibility for the groove while the saxophone plays its solo part on it. Equally to MSA the saxophone has a perfect sync on its side and is transmitted back with the roundtrip time but in comparison to MSA this has no disturbing effect on the rhythm instrument in LBA. The saxophone is delayed by the roundtrip delay time, which adds an artificial laid back style on it and hence this playing constellation is not to be considered as problematic anymore. LBA of course does not work for unison music parts in which both parties have to play exactly on the same beat and at the same time. The perceived roundtrip delay on the master's end ranges between 50 ms up to a maximum of 100 ms but still depends on the musician's subjective perception and the bpm (beats per minute) of the

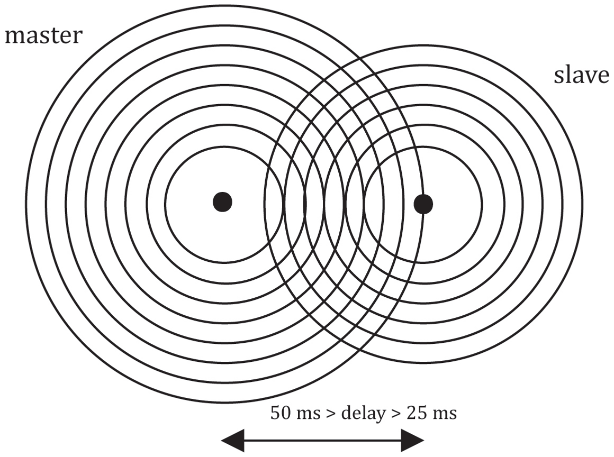


Figure 6: Laid back approach

actual song [7]. Figure 6 equals the MSA principle but due a maximal one-way delay of 50 ms and the determination of a rhythm and a solo section this situation leads to an artificial “laid-back“ effect.

LBA is used when the delay ranges in areas slightly beyond the 25 ms RIA threshold. Again SoundWIRE and Soundjack represent potential candidates, beside the Musigy [2] software as one of the fist commercial products. It provides audio delays, which range at the edge between RIA and LBA.

3.4 Delayed Feedback Approach (DFA)

In case the 25 ms delay threshold is exceeded, DFA tries to make musicians feel like playing with the RIA by delaying the player’s own signal artificially: By principle delays beyond 25 ms lead to either LBA or MSA styles in which the master hears the slave with a delay equal to the roundtrip time while the slave plays in perfect sync. When delaying the playback of the master’s signal, both sounds finally have a closer proximity at the master’s ear, which improves the problematic delay gap in MSA or reduces the laid back effect in LBA. The larger the self-delay the better the synchronization of both signals. The best synchronization can be reached with a self-delay equal to the roundtrip-time. This principle is illustrated in figure 7. Though DFA improves the delay situation between two musicians, it is no doubt that a delay of one’s own signal typically can be considered as inconvenient and not natural. The larger the delay gets and the louder the instrument’s direct noise, the worse the realistic instrument feel and playing conditions. This is especially valid for any acoustic instrument such as a violin or drums. On the other hand DFA can be a suitable approach for the synchronization of remote playback sound sources. In case of e.g. two DJ’s turntables are connected with each other, a delay of the turntable’s output would not lead to timing-problems. Unlike human beings a machine’s playback behavior does not depend on an inner time or feel and hence can

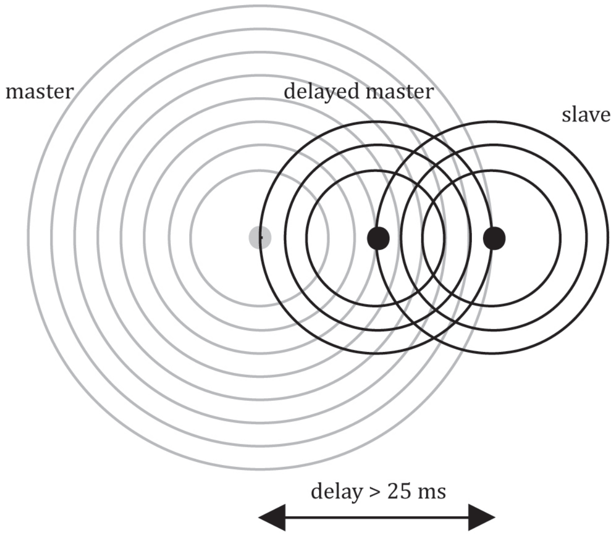


Figure 7: Delayed feedback approach

of course reproduce delayed sounds without loosing any kind of rhythm. Systems based on DFA are eJamming [1] and the NcMP project of a research group (Network-centric Music Performance) at the University of Braunschweig, Germany [13].

4 Conclusion and Future Work

Apart from audio engineering, network and music skills, the awareness of delay dimensions and their musical consequences is the main requirement for a successful network music performance. This workshop firstly describes the main technical aspects of the Soundjack software, in order to figure the audio- and network latency between two musicians. Secondly, it explains the related categories of delay influencing musical interaction. Depending on the actual network connection and the respective delays, the user can now consciously apply the suitable category of musical interplay, which allows him to perform under any given network situation. In parallel this gives him awareness of actual possibilities and limitations in his current situation. However, due to the high amount of interdisciplinary knowledge, distributed music has so far been used by a small community of experts in IT as well as in music. Hence it cannot be considered as a major technology for musical interaction, yet. Despite the existence of first commercial products, musicians and sound engineers remain passively in terms of accepting and applying this new approach. As the technical facts clearly prove the feasibility of network music performances, we hope to motivate musicians and engineers with this workshop to take advantage of the musical possibilities distributed music can offer.

In the future we will further investigate in the realistic interaction approach for the Internet in order to finally increase the radius, in which RIA can be applied. As a new field of interest we will examine the delay and interaction restrictions for conducted orchestrated music with the final goal of developing a decent and versatile low delay audio and video streaming solution.

References

[1] ejamming website: www.ejamming.com, August 2008.

[2] Musigy website: www.musigy.com, August 2008.

[3] Skype website: www.skype.com, August 2008.

[4] Alexander Carôt. Livemusic on the internet. In *Diplomarbeit*, Fachhochschule Lübeck, 2004.

[5] Alexander Carôt, Ulrich Krämer, and Gerald Schuller. Network music performance in narrow band networks. In *Proceedings of the 120th AES convention*, Paris, France, May 2006.

[6] Alexander Carôt, Alain Renaud, and Bruno Verbrugghe. Network music performance with soundjack. In *Proceedings of the 6th NIME Conference*, Paris,France, June 2006.

[7] Alexander Carôt and Christian Werner. Network music performance – problems, approaches and perspectives. In *Proceedings of the "Music in the Global Village" - Conference*, Budapest,Hungary, September 2007.

[8] Alexander Carôt, Christian Werner, and Alain Renaud. Audible icmp echo responses for monitoring ultra low delayed audio streams. In *Proceedings of the 124th AES-Convention*, Amsterdam, the Netherlands, May 2008.

[9] C. Chafe, M. Gurevich, G. Leslie, and S. Tyan. Effect of time delay on ensemble accuracy. In *Proceedings of the International Symposium on Musical Acoustics*, Nara,Japan, March 2004.

[10] C. Chafe, S. Wilson, R. Leistikow, D. Chisholm, and G. Scavone. A simplified approach to high quality music and sound over ip. In *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00)*, Verona,Italy, December 2000.

[11] Martin P. Clark. *ATM Networks – Principles and Use.* Wiley-Teubnerl, 1996.

[12] K. Egevang and P. Francis. RFC 1631: The IP network address translator (NAT), May 1994. Status: INFORMATIONAL.

[13] Xiaoyuan G, Matthias Dick, Ulf Noyer, and Lars Wolf. Nmp – a new networked music performance system. In *Proceedings of the 4th NIME Conference*, June 2004.

[14] Richard T. Griffiths. History of the internet, internet for historians : www.let.leidenuniv.nl, August 2008.

[15] International Telecommunication Union (ITU). The status of voice over internet protocol (voip) worldwide, 2006. In *The Future of Voice*, 2007.

[16] A. Kos, B. Klepec, and S. Tomazic. Techniques for performance improvement of voip applications. In *Proceedings of the 11th Electrotechnical Conference MELECON*, 2002.

[17] Marshal Mc Luhan, Quentin Fiore, and Jerome Agel. *The Medium is the Massage: An Inventory of Effects*. Gingko Press, 1996.

[18] J. Postel. RFC 768: User datagram protocol, August 1980.

[19] J. Postel. RFC 791: Internet Protocol, September 1981.

[20] Nathan Schuett. The effect of latency on ensemble performance. In *Bachelor Thesis*, CCRMA Department of Music, Stanford University, 2002.

[21] William Stallings. *ISDN and Broadband ISDN with Frame Relay and ATM*. Prentice-Hall, third edition, 1995.

[22] J.M. Valin, T.B. Terriberry, C. Montgomery, and Gregory Maxwell. A high-quality speech and audio codec with less than 10 ms delay.

[23] James R. Wilcox. *Videoconferencing – the whole picture*. Telecom Books, third edition, 2000.