

PLAYING THE NETWORK: THE USE OF TIME DELAYS AS MUSICAL DEVICES

Juan-Pablo Cáceres

Center for Computer Research in
Music and Acoustics (CCRMA)
Stanford University

jcaceres@ccrma.stanford.edu

*Alain B. Renaud**

Sonic Arts Research Centre (SARC)
Queen's University Belfast
Northern Ireland

alain.renaud@qub.ac.uk

ABSTRACT

The native time delay of audio transmission over high speed networks is used to create musical devices to play *on* and *with* the network. Different strategies that have been part of the practice of the *Net vs. Net* collective—a permanent network ensemble playing exclusively over wide area digital networks—are discussed. A tool to synchronize displaced musicians visually cue the performers and the audience. It also activates synchronized musical processes. The latency is used to create delay effects and reverbs embedded on the bidirectional path using feedback delay network filters. We also present a technique to play distributed rhythmic patterns over the network that are designed to sound different—with diverging rhythmic and phase structures—on each location.

1. INTRODUCTION

Recent developments in networking technology make meaningful and creative networked musical interactions a reality. Now that the infrastructure is in place and reliably working, musicians connected over the network have the opportunity to concentrate on creating new forms of musical expression that use the network as an integral part in the creation and performance. In this specific case, the network is considered as a medium which brings its own acoustic contribution to the music making process rather than being a link between two spaces. This approach takes into account particularities introduced by the network such as latency and the combination of virtual and physical acoustic spaces as a set of devices to develop musical material that, in some cases, couldn't be created outside of a networked situation.

This new potential paves the way for developing strategies to use the network creatively and makes music playing in the network rather than over the network a reality. They include:

- Using networked time delay in a musical fashion rather than constantly trying to counter it.
- Combining the acoustics of physical spaces with the acoustics brought by the network itself.

- Translating the various latencies into spatial diffusion algorithms, making the distance “audible.”
- Taking advantage of the scattering of musicians over the network to create a de-multiplication of musical patterns affected by the various network latencies.
- Exploiting the phenomena of synchronization and asynchronization leading to the creation of material that sounds different on each end.

In order to elaborate a consistent methodology to apply these techniques in a real-time networked scenario, the two authors have created a permanent networked ensemble, the *Net vs. Net* collective [5].

2. NET VS. NET COLLECTIVE

The ensemble was started as a result of several years of musical collaboration between the authors over the network. The two original members decided to initiate the collective as a way to demonstrate the potential of digital wide area networks for meaningful networked musical interactions. The first piece created by *Net vs. Net* was performed at the annual CCRMA concert (Fig. 1) in November 2007 [1]. It combined several strategies presented later in more details, including visual cues for synchronization and a multi-channel acoustic spatialization of signals based on the conditions of network delays. It is important to point out that *Net vs. Net* is an example of musical practice led research in the emerging field of network music performance (NMP). The collective doesn't attempt to propose a robust framework but rather relies on its experiences over the network to develop unique strategies to make network music playing as interactive and innovative as possible.

3. DISTANCE AND SPACES

Trying to deal with the time delay introduced by network latency has been the main focus for many researchers, including ourselves, in the NMP area. Currently, the main approach to counter the delay is to compromise, most of

*Co-first author

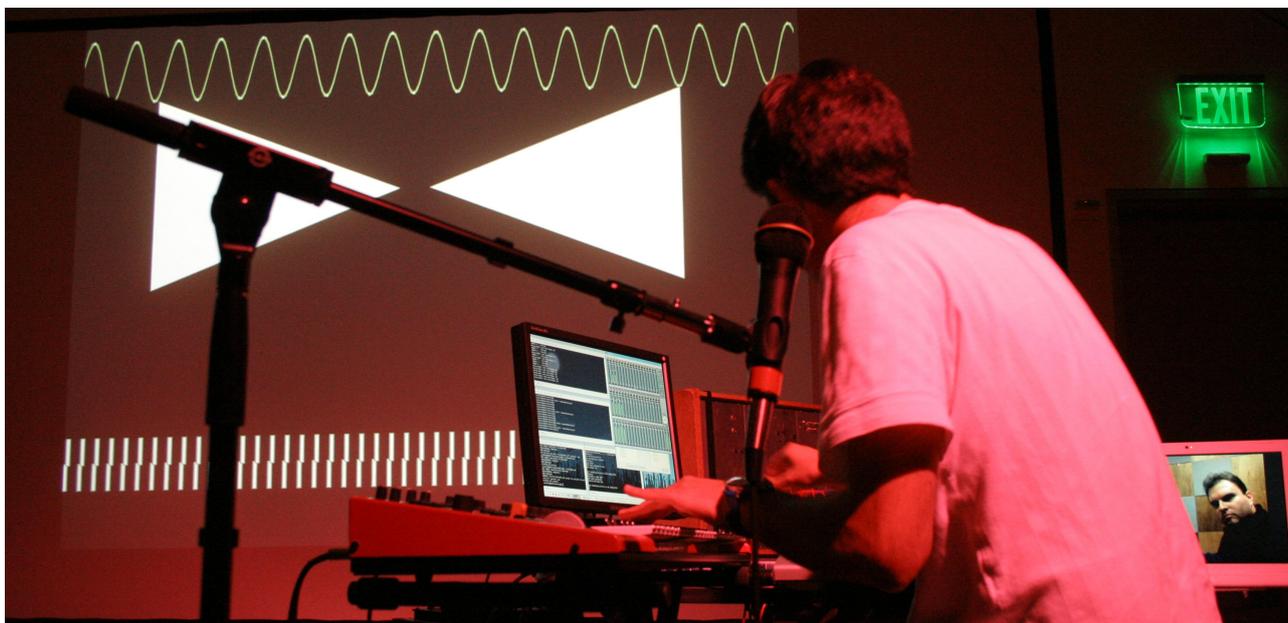


Figure 1. *Net vs. Net* in action, CCRMA Concert 2007: the screen shows visual cues used for synchronization of musicians and identification of performers for the audience.

the time by artificially adding delay to one of the participants in the performance, so that the amount of delay can translate into musical terms.

For example, Goto et al. proposed an approach to synchronize musicians over wide area networks with long delays [15, 14]. The connection delay is increased by a constant corresponding to a repetitive musical structure, e.g., a 12-bar blues chord progression. Each musician is then playing on top of a delayed version of all the others.

Even though this works to synchronize musicians, it imposes constraints on the musicality of the ensemble, and from the performer’s point of view it feels almost like playing on top of a recording, since the musician actually plays on a signal that was recorded a few seconds ago. NINJAM [6], a software application that implements this technique provides several audio examples on its website.

Other approaches include the use of shared virtual acoustical spaces [9], where several microphones and a dummy head are used to generate a virtual location mixing real halls. Audio prediction techniques [19, 18] have been implemented using messaging rather than audio for networked tabla performances.

The musical explorations of *Net vs. Net* attempt to deal with the network conditions “as they are.” The goal is to immerse ourselves and play in the network by considering it a space in itself rather than using the network to just connect two or more sites together. We explore different techniques that use the delay as a way to create rhythmic structures and spatial behaviors that make the network distance “audible.” We go as far as re-naming the network musicians to “nodes” where each node interacts with each other based on its location on the network mesh. When two or more nodes are involved, the combination of acoustic spaces and the latencies between nodes creates

secondary virtual spaces which translate into additional reverbs or echoes.

4. NODE TO NODE CONNECTIVITY: THE QUEST FOR SYNCHRONOUS INTERACTIONS

One of the challenges encountered with *Net vs. Net* is the ability to transmit crucial musical cues over the network in a synchronous fashion. Even if the distance is unequal between nodes and even if we propose to use the unevenness for artistic purpose, it is important that each node responds simultaneously to cues, at least unidirectionally. We approached the issue of synchronicity by developing a Master Cue Generator (MCG) which broadcasts various messages to each node from a central location.

The application, currently implemented with Max/Msp [4] is able to analyze the latency between the master and the nodes and compensate the delay to ensure that each cue arrives simultaneously to each node. The messages are formatted as standard OpenSoundControl (OSC) messages [21, 20] which allows them to be parsed by any receiving system capable of dealing with OSC.

The MCG also broadcasts important musical information by providing a basic structure to the nodes playing over the network, such as which section of the piece the nodes are in as well as warning messages that the piece is about to switch to another section. The MCG is fully customizable and new messages can be created for specific pieces or structured improvisations. There are two distinct types of cues:

Passive cues Cues that are sent out as a piece of information to the nodes but don’t actually automatically affect the sound; examples include suggestions that

one node should decrease its amplitude or a flashing warning that the piece is about to switch to another cue.

Active cues Cues that actually trigger elements on the nodes in an automated fashion; examples include the spatialization of the sound sources in remote spaces or the triggering of a remote oscillator.

Figure 2 shows the MCG sending passive and active messages to unevenly scattered nodes around the network. A feedback mechanism from the nodes is used so that the MCG can estimate the latency for each node. This is achieved by having the MCG broadcast a trigger message (T) to the nodes whenever an important cue is about to happen in a piece. Once this information is received by the MCG, it can automatically compensate to reach each node simultaneously. The path from the MCG to each node is represented by curved arrows with associated letters. In the case below, the MCG will take the longest path (B in this case) and add delay so that the shorter paths all equal path B. This doesn't prevent the nodes from using the combination of network delays creatively; only active and/or passive cues are transmitted to synchronize the ensemble on a per-event basis rather than on a continuous basis.

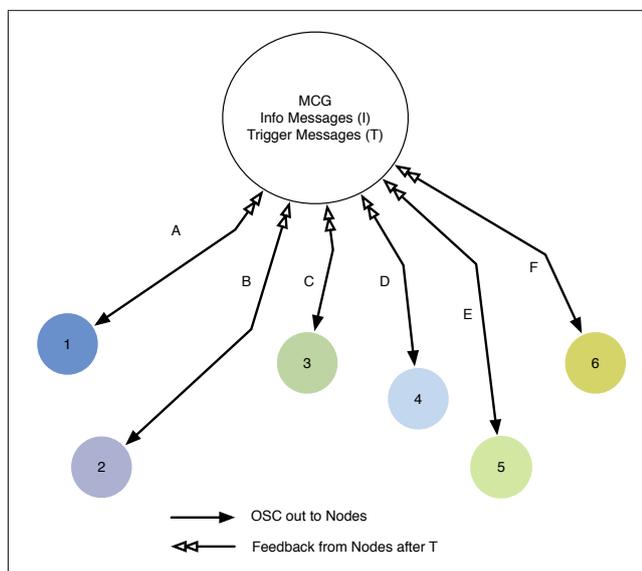


Figure 2. The relationship between the MCG and unevenly scattered nodes

Once the MCG to nodes relationship is established, the actual cue broadcasting distribution acts as if the MCG is virtually located in the center of the network at equal distance to each node.

In parallel with the MCG, the JackTrip [10] high quality audio streaming application is used to send and receive uncompressed audio signals to and from each node participating in the performance. The performances that we have designed so far have used up to eight channels of audio over the network to allow multi-channel diffusion on each node.

5. VISUALIZATION

A visualization tool has been developed using Processing [7]. The graphical tool, which receives OSC messages from the MCG, is currently able to receive three types of cues: time cues, interaction cues and spectral cues (Fig. 1).

Time cues These are the most important types of cues to keep the ensemble together. They are synchronized with the audio triggering cues. Time cues are broadcast to the nodes through the MCG, which also acts as a time referential for the entire ensemble. The latency to the nodes are constantly calculated—a good estimate is obtained by pitch-tracking a *Sound Ping* as in Chafe and Leistikof [12]—and compensated for by adding a proportional amount of delay to each node so that the cue arrive at destination at exactly the same time.

Interaction cues These are directly mapped to the audio signal generated by each node and are concentrated into a graphical canvas displaying various geometrical forms. Each form is dynamically modified in real time depending on the type of signal generated by the node. Each form has a mathematical relationship to, for example, allow one node to influence the relationship with the other node. This offers an important aspect often ignored in networked music performance: the identification of the interplay between each specific node.

Spectral cues These represent graphically the overall signal generated by all the nodes. They provide an overall picture of all the interactions taking place over the network. They are currently represented as a succession of sine waves that are changing in real time depending on the number of nodes present on the network and the overall dynamic range of the signal.

The visualization tool has brought a good level of interactivity in our various performances. It allows the performers as well as the public to better identify visually the sonic material generated by the participants on the network.

6. FEEDBACK DELAY AND REVERBERATION

With short delays the network can be used as a reverberating medium [11]. On longer delays like the ones between San Francisco and Belfast (~130 milliseconds, one-way), the feedback delay is no longer perceived as reverberation but as an echo. We use this technique in order to “bounce” sounds from one location (ipsilateral) to a second one (contralateral). This technique not only uses the network delay as the memory part of a filter, but also has the advantage of making the delay—and consequently the sound distance—audible to the performers.

The implementation uses a similar technique like the one described by [Chafe](#), but instead of a Schroeder-style reverberator [16] we use a multichannel feedback delay network (FDN) [17] in the feedback comb filter sections.

6.1. General FDN

A general FDN of order 4 is illustrated on Figure 3. The time-update for this FDN can be written as:

$$\begin{bmatrix} x_1(n) \\ x_2(n) \\ x_3(n) \\ x_4(n) \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \end{bmatrix} \cdot \mathbf{I} \cdot \mathbf{q} \cdot \begin{bmatrix} x_1(n - M_1) \\ x_2(n - M_2) \\ x_3(n - M_3) \\ x_4(n - M_4) \end{bmatrix} + \begin{bmatrix} u_1(n) \\ u_2(n) \\ u_3(n) \\ u_4(n) \end{bmatrix} \quad (1)$$

Where \mathbf{I} is the identity matrix and

$$\mathbf{q} = \begin{bmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & q_{34} \\ q_{41} & q_{42} & q_{43} & q_{44} \end{bmatrix} \quad (2)$$

The outputs are given by

$$\begin{bmatrix} y_1(n) \\ y_2(n) \\ y_3(n) \\ y_4(n) \end{bmatrix} = \begin{bmatrix} x_1(n - M_1) \\ x_2(n - M_2) \\ x_3(n - M_3) \\ x_4(n - M_4) \end{bmatrix} \quad (3)$$

The ipsilateral and contralateral sides can both act as inputs and outputs on this structure, as indicated on Figure 3. The scheme is therefore symmetric and bidirectional.

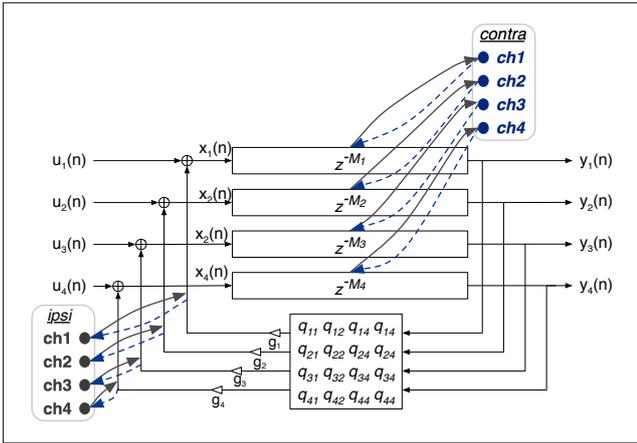


Figure 3. Order 4 Feedback Delay Network (FDN) showing inputs and outputs of the ipsilateral and contralateral nodes

6.2. Network Implementation

In the network each channel is implemented as a comb filter, as illustrated on Figure 4. In this case the configuration is symmetric, i.e., both the ipsilateral and contralateral side can read and write on its ends and use the comb filter as a “bouncing” device.

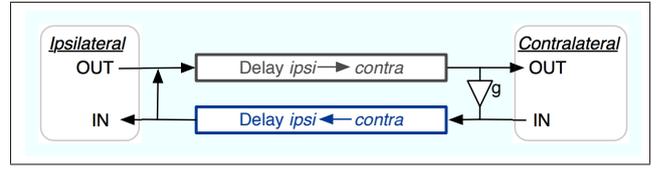


Figure 4. Distributed Comb Filter used on each channel

Depending on the latency, the FDN topology can be used as a reverberator or as a delay type of effect. We currently use the latter in *Net vs. Net* since most of our musical practice is transcontinental—with delays that are long enough to be perceived as echoes.

The simpler configuration of the FDN as a delay effect occurs when $\mathbf{q} = \mathbf{I}$ (identity matrix). The configuration in this case includes four independent comb filters (Fig. 5a). The performer has control over the decaying rate (g) for each channel and can also extend the base delay line length (the one that is fixed by the network conditions) by some arbitrary time.

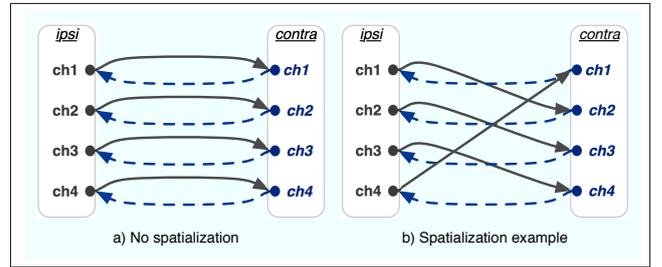


Figure 5. Two configurations for the network FDN

A further extension consists of dynamically changing the matrix \mathbf{q} to generate specific spatial and rhythmic effects. For example, \mathbf{q} can be transformed as:

$$\mathbf{q} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \quad (4)$$

On each row, there is a “1” that moves to the next column. This is equivalent to panning the sound between speakers, therefore panning curves can be used to dynamically change the matrix from one permutation to the next. Figure 5b shows the configuration after the permutation is completed. This particular example shows how the delay effect is not (spatially) static anymore. The sound moves from one channel/speaker to the next with the same rhythmic structure of the echoes, coupling space and audio through the network delay.

Controlling the decaying rate (g) base delay line length on each channel, the performer can generate different patterns with predictable or more chaotic characteristics. The “1” in the matrix can also continue to move on each row. This dynamically changes the spatial configuration of the echoes.

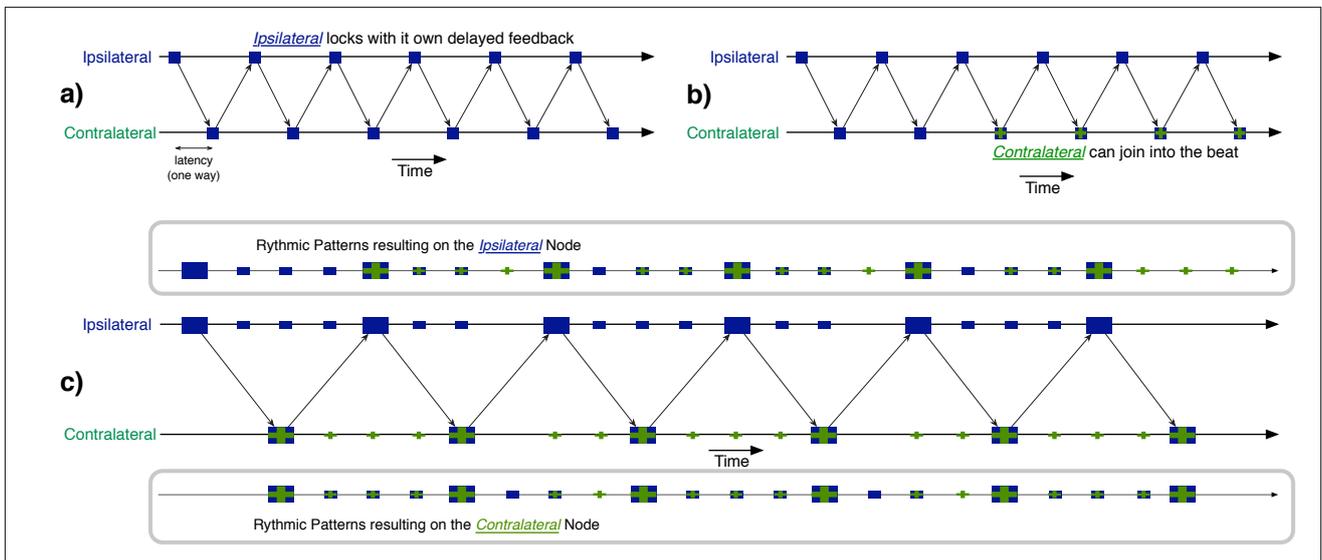


Figure 6. Feedback Locking: a) A common pulse is distributed to both locations. b) Synchronization of a remote performer (contralateral) to the pulse originating at the ipsilateral node. c) Example of resulting rhythmic combinations.

7. DISTRIBUTED RHYTHMIC PATTERNS OVER THE NETWORK

At a local level (small time frames) and depending on the delay, each node on a network mesh hears different musics. Sound events arrive at slightly (or not so slightly if the delay is long) different times on each location. One technique to deal with this problem is to compensate for the differences on transmission delays, but this also requires that the musicians hear themselves with a constant delay [8].

We use this fact to explore scenarios in which rhythmic patterns are explicitly designed to sound different on each location. Each node broadcasts its audio material to the others, and depending on the network situation, different patterns are audible on each location. The synchronization is implemented using the MCG outlined in Section 4. This creates, depending on the distance between nodes and its number, simple or more complex sequences.

7.1. Feedback Locking

One technique which works well to play music over a network is to follow and cue in to a pulse. We call this method *Feedback Locking*. A local sound is *fed-back* once (or more) to the original location after it reaches the remote node. The musician can then synchronizes with its own sound, much like guitar players synchronize with delay pedal effects. The advantage of this approach is that a second performer—located on the other end of the network—can then synchronize to this beat. In this scenario, the performance follows a common pulse, but what happens on each location is out of phase in absolute time terms.

Figure 6 shows the feedback locking process for two remote nodes. Each axis of the diagram represents one of the two nodes—the ipsilateral and the contralateral. The sounds generated at the ipsilateral end are represented by a

square, while a cross represents the ones generated at the contralateral node. The delay is assumed to be constant (no jitter) and symmetric between the two locations. This is a very good estimate for performances using the present state of the art high-speed networks, such as Internet2 [3] and Geant2 [2].

The vertical distance between the axis is a representation of the latency between the two locations. This time delay is also labeled at the bottom of Figure 6a, and represents the time for a sound to go from the ipsilateral to the contralateral node. It takes the same time for a sound to travel in the opposite direction, assuming that the path is symmetric.

Figure 6a shows the ipsilateral node synchronizing with its own feedback. In Figure 6b, the contralateral node synchronizes to its local version of the pulse generated by ipsilateral node.

The distributed pulse can be use to play rhythmic music. Performing patterns that follows this beat result in mixtures that are different on each node. On Figure 6c an example is presented. The rhythmic combination on each location differs significantly, while at the same time both nodes follow the distributed, feedback-locked, pulse. This allows network-only musical possibilities in which *one* performance generates *two* (or even more, depending on the number of nodes) musical results.

7.2. Bi-located Patterns

It is possible to extend the technique presented in Section 7.1 to different types of patterns. A simple example can serve to illustrate its applications and possible implications. Figure 7 shows four cases of different patterns that are generated in the two locations (ipsilateral and contralateral). This illustrates how, depending on the latency and the pattern played, the music heard at the ipsilateral node relates to the one heard at the contralateral side:

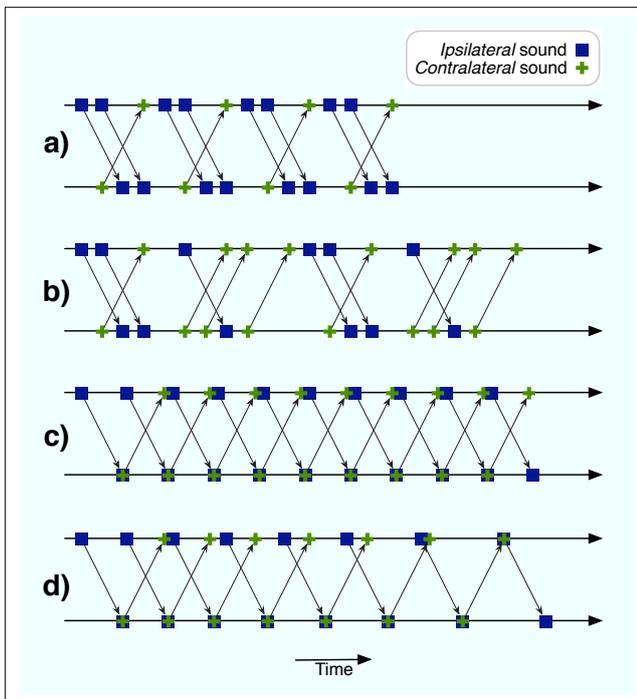


Figure 7. Examples of distributed pattern interactions

- a) **Dislocated identical pattern** This case shows an example where the same pattern is heard on both locations, but the sounds that compose it are part of the previous pattern on the other node. Even if they sound the same, they are actually interlocked.
- b) **Different patterns** For longer sequences this example shows how the production on each node differs. For example, on the contralateral side there are always three or four sounds heard in sequence, however on the ipsilateral, isolated and two sounds are heard more frequently. This happens when the period of each pattern is different.
- c) **Constant pulse** This simpler examples shows how for a constant and identical pulse on each location, the contralateral hears the two sounds totally synchronized while they are slightly out of phase on the ipsilateral. This is what happens when musicians try to play together and what causes the tempo deceleration [13].
- d) **Ritardando and phasing** The ipsilateral node plays a pulse with a regular slowing in tempo, while the contralateral synchronizes perfectly to it on its location. The effect here is that while on the contralateral node the sounds are always played in unison, the ipsilateral hears a constant shift (phasing) in the relation between each sound.

For situations where the time delay is on the reverberation range, example d) suggests how the phasing can be applied to the sound quality instead. This would make the sounds itself different on each location and it is something that needs further investigation.

7.3. Multiple nodes

A logical extension to the distributed patterns technique is the addition of more nodes. It is expected that more complicated interactions and patterns will occur. A consistent framework to understand it is still to be developed.

8. CONCLUSIONS AND FUTURE MUSIC

Broadband Internet technologies are developed enough to support real-time high quality audio transmission. Everything indicates that these technologies are going to become ubiquitous in the near future beyond National Research and Education Networks (NRENs). The field of NMP is also growing at a steady pace with an increasing number of artists and music technologists becoming interested in the subject. We hope that the musical strategies presented in this paper to play *on* and *with* the network using the inherent delay as part of the musical structure will encourage further research and musical practice that goes beyond jamming sessions and the performance of existing repertoire on this new medium.

We intend to develop further the distributed patterns technique to include a larger number of nodes. We also plan to expand the *Net vs. Net* collective and include acoustic instrument players in a consistent way. This will also allow us to go further in the exploration of the sharing of acoustic spaces, especially when the combination of physical (the acoustics of a room) and virtual (the network) ones are combined. The challenge is to incorporate acoustic and electronic instrument in a way that makes sense in this new medium opening in front of us.

References

- [1] CCRMA Concert Series, 2008. URL http://ccrma.stanford.edu/concerts/c_schedule.html.
- [2] Geant2, 2008. URL <http://www.geant2.net/>.
- [3] Internet2, 2008. URL <http://www.internet2.edu/>.
- [4] Max/MSP, 2008. URL <http://www.cycling74.com/products/maxmsp>.
- [5] Net vs. Net, 2008. URL <http://www.netvsnet.com/>.
- [6] NINJAM: Realtime music collaboration software, 2008. URL <http://www.ninjam.com/>.
- [7] Processing, 2008. URL <http://processing.org/>.
- [8] N. Bouillot. The auditory consistency in distributed music performance: a conductor based synchronization. *ISDM (Information Science for Decision Making)*, (13):129–137, 2004.

- [9] J. Braasch, D. L. Valente, and N. Peters. Sharing acoustic spaces over telepresence using virtual microphone control. In *Proceedings of the 123th Convention of the Audio Engineering Society*, Oct. 2007.
- [10] J.-P. Cáceres. Jacktrip: Multimachine jam sessions over the Internet2, 2008. URL <http://ccrma.stanford.edu/groups/soundwire/software/jacktrip/>.
- [11] C. Chafe. Distributed internet reverberation for audio collaboration. In *Proceedings of the AES 24th International Conference*, 2003.
- [12] C. Chafe and R. Leistikof. Levels of temporal resolution in sonification of network performance. In *Proceedings of the 2001 International Conference on Auditory Display*. ICAD, 2001.
- [13] 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*, 2004.
- [14] M. Goto and R. Neyama. Open remoteGIG: An open-to-the-public distributed session system overcoming network latency. *Transactions of Information Processing Society of Japan*, 43(2):299–309, 2002. (in Japanese).
- [15] M. Goto, R. Neyama, and Y. Muraoka. RMCP: Remote music control protocol - design and applications. In *Proceedings of International Computer Music Conference*, pages 446–449, Thessaloniki, Greece, 1997.
- [16] J. A. Moorer. About this reverberation business. *Computer Music Journal*, 3:13–28, June 1979.
- [17] D. Rocchesso and J. Smith. Circulant and elliptic feedback delay networks for artificial reverberation. *Speech and Audio Processing, IEEE Transactions on*, 5:51–63, 1997.
- [18] M. Sarkar. Tablanet: a real-time online musical collaboration system for indian percussion. Master’s thesis, MIT Media Lab, Aug. 2007.
- [19] M. Sarkar and B. Vercoe. Recognition and prediction in a network music performance system for Indian percussion. In *NIME '07: Proceedings of the 7th international conference on New Interfaces for Musical Expression*, pages 317–320, New York, NY, USA, 2007. ACM.
- [20] M. Wright. Open sound control 1.0 specification, 2002. URL http://opensoundcontrol.org/spec-1_0.
- [21] M. Wright. Open sound control: an enabling technology for musical networking. *Organised Sound*, 10:193–200, 2005.