THE INFLUENCE OF PERCEPTUAL ATTACK TIMES IN NETWORKED MUSIC PERFORMANCE

ÁLVARO BARBOSA[1], AND JOÃO CORDEIRO[1]

[1] *Research Center for Science and Technology of the Arts (CITAR), Catholic University - Oporto, Portugal*
{abarbosa;jccordeiro}@porto.ucp.pt

It is well known that Latency has a highly disrupting influence in music performance, affecting musicians individually on their own acoustic feedback and collectively at the overall musical response. In musical communication over the Internet one of the most promising approaches, to minimize the constraints caused by extreme latencies, is to adapt the musical practices to perceptual audio features that influence the latency threshold tolerated by musicians. According to the information extracted from these descriptors, it is possible to shape the musical language at the compositional stage or at the performative level. Research has previously been conducted exploring the relationship between this threshold and features such as the musical Tempo and Loudness. Nevertheless, further studies remains to be done, particularly concerning the ability to synchronize delayed performances for different expressive musical qualities related with performing technique at the level of individual notes. This paper presents a pilot study that suggests a better tolerance to communication latency, for musical pieces based on notes performed with slow attack sounds.

INTRODUCTION

With the advent of Internet computing and the possibility of ubiquitous long distance acoustic communication, the prospect of geographically displaced musical performance became available to a worldwide community.

However, in global networks communication Latency (net-delay) is extreme and causes a highly disrupting effect in remote musical performances, especially in more traditional musical forms, which are mostly driven by rhythm and melody involving very tight synchronicity, to achieve a desirable real-time awareness of musical acoustics.

In fact, requirements for networked musical performance imply latency thresholds that are amongst the most demanding in Computer Supported Collaborative Work (CSCW), providing a leading framework of critical research for this particular field of Computer Science.

1 LATENCY THRESHOLDS AS A FUNCTION OF MUSICAL FETURES

While in a **vocal conversation** it is possible to maintain it even with one-way delays of up to **500 ms** [1], when trying to perform any sort of rhythmical piece on gestural based musical instruments, the communication delay threshold, that allows musicians to maintain a synchronized and smooth **musical interaction**, reduces drastically to the order of **tens of milliseconds**, as reported on a number of experimental studies conducted since 2002: [2] [3] [4] [5] [6] [7] [8].

These studies, attempt to determine the **Ensemble Performance Threshold (EPT)**, "*the level of delay at which effective real-time musical collaboration shifts from possible to impossible*" [2], and consistently the results fall within the range 20 to 40 ms.

Yet, the EPT is usually determined based on experimental setups that focus on latency variations, maintaining fixed values in other musical features. Moreover, in each study the EPT has been determined for particular performative conditions, without introducing significant variations in other musical facets, such as:

- The music expressive qualities (Dynamics and Articulation)
- The music style (rhythm, melody, harmony)
- The music perceptual qualities (pitch, texture, timbre)
- The music structure/form
- The musician's experience and practicing strategies
- Complementary feed-back modalities (visual, tactile)
- The listening conditions

In this sense, further research is necessary in order to determine if the established EPT range (20 to 40 ms) changes according to different perceptual audio features, which are fundamental to the music performance.

In particular, this paper presents a study on the influence in EPT of very specific audio feature - **attack time** -. This results builds upon recent studies by the authors, and other researchers, that have explored in EPT of related features such as musical **Tempo** and **Loudness/Dynamics**.

1.1 Latency adaptive Tempo

The relationship between musical Tempo and the amount of latency, performers can tolerate in order to maintain a smooth music interaction, is often empirically perceived in EPT experiments. In these experiments it is common that subjects start to slow down the Tempo when the EPT is reached and just before they break the performance apart. This impulse of slowing down the Tempo in the presence of latency motivated an early experiment conducted in 2005 by Álvaro Barbosa and Alexander Carôt.

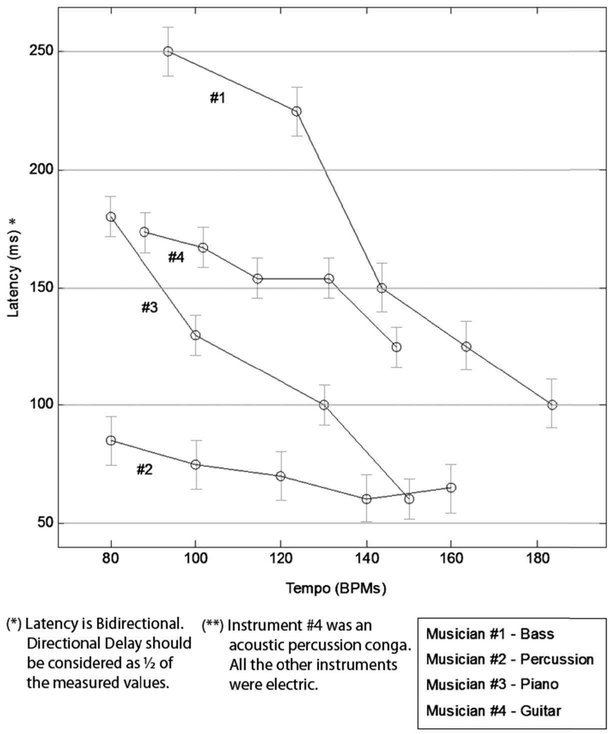

Figure 1: Performing with Self-Delay/different Tempos from the Barbosa/Carôt experiment (2005)

In the experiment, simulated network latency conditions were applied to the performance of four different musicians playing jazz standard tunes with four different instruments (Bass, Percussion, Piano and Guitar). The objective consisted in determining the maximum individual latency tolerance applied to the auditory feedback from the musician's own instrument.

For this purpose a studio system was set up, so that performers would listen to the feedback from their own instruments through headphones with variable delays. Performances were synchronized with a metronome over several takes with different beats per minute (BPM). For each take the feedback delay was increased until the musician wasn't able to keep up a synchronous performance.

In figure number 1 is presented a graphic with the results from this experiment. It is clear that regardless of instrumental skills or music instrument, all musicians were able to tolerate more feedback delay for slower Tempos.

The only exception, to the decreasing tendency of these curves, occurs for the percussionist when raising to 160 BPMs, which is related to a synchronous overlap over the music rhythmic structure (from 80 BPM to 160 BPM) and the fact that with acoustic percussion instruments it is are very hard to totally isolate the performer from direct instrument sound. Therefore, it is reasonable to assume that there is an **inverse relationship between Musical Tempo and Latency-Tolerance**.

This direct dependency between musical Tempo and latency tolerance can be regarded as a more general concept – Latency Adaptive Tempo (LAT).

The basic application principle of LAT consists of a simple function, for audio network communication software systems, that dynamically adapts the Musical Tempo (typically a referenced by a metronome sound) to the maximum value tolerated by the least "latency-tolerant" musician of a network ensemble. This dynamic adaptation is based on real-time latency measurement between peers.

Input variables of this function are musicians' profiles and latency value at a given moment. The output of the LAT function will be the Tempo value (typically in BPMs) that is less disrupting for the group musical practice.

LAT allows musicians to rehearse music as fast (in terms of Musical Tempo) as their Network connectivity speed allows them to.

This concept was implemented into the Public Sound Objects (PSOs) system [9], with respective adjustments to the Musical Tempo concept and latency-tolerance requirements in this particular Networked Music instrument.

1.2 Latency Adaptive Loudness

In a real world acoustic environment, the perception of sound loudness decreases naturally when the distance between the receiver and the sound source increases. The further away is the position of the sound source, the lower is the perception of loudness and the higher is the acoustic propagation latency.

This real-world inverse proportion between sound loudness and communication latency inspired the idea of a network music instrument, which adapts to the Internet latency and was implemented in 2003 by Jörg Stelken in the Software *PeerSynth* [10]. *PeerSynth* is a per-to-peer sound synthesizer, which supports multiple users displaced over the internet, measuring the latency between each active connection and adaptively lowering the incoming sound volume of each user's contribution to the overall soundscape, proportionally to the amount of delay measured in his connection. A similar approach was followed in the *AALIVENET* System [11].

Loudness and softness of music is usually referred to as Dynamics. However, musical dynamics is not only regarded as the sound volume of the overall musical piece, but also as the articulation between individual musical notes (staccato, legato etc.), as well as the performing techniques applied to emphasize each note (particularly on its attack and decay). In this context, the perceptual attack time of individual notes is a key feature.

2 PERCEPTUAL ATTACK TIME (PAT)

Like Tempo and loudness, individual note interpretation can have a significant impact in the range of values obtained for the EPT. In fact, the attack of a musical note is a critical feature in discerning the perception of simultaneity for overlapping sound event.

As part of the experiments conducted by Al Bregman in the late 90's on Auditory Scene Analysis [12] there is a particular section on the perception of order in overlapping sound events for different rise times. Bregman demonstrates that "*the auditory system seems to be particularly interested in sounds with abrupt onsets and that such sounds stand out better from a background of other sounds than do slow-rising sounds*". This experiment[1] clearly reveals that it is harder to perceive the order of overlapping sound events when their rise time (attack) is slower. In other words, overlapping sounds with slower rise times are better perceived as synchronous even when their onsets are not physically simultaneous (in this experiment there was a time delay of 120 ms between the sound events).

While Bregman is not referring specifically to Perceptual Attack Time (PAT), a difficulty in discerning the order of sounds will obviously lead to a difficulty in discerning whether or not two sounds are perceptually simultaneous.

However, the fact that two sounds are physically synchronous does not necessarily make them sound perceptually synchronous.

PAT is a subjective measure of the time that is perceived by the listener as the moment of rhythmic placement for a musical sound event. According to Matt Wright [13] "*for highly percussive sounds, the PAT might be the same as, or just a few milliseconds after, the onset time, but for sounds with a slow attack, for example, bowed violin, the PAT might be dozens of milliseconds after the onset*".

Although, Bregman's studies or the PAT definition are not used specifically in reference to the mutual awareness of sounds being produced in a bilateral music performance, it can be argued that a difficulty in discerning the order of sounds events, will inevitably lead to conditions in which is hard for instrumental performers to maintain a synchronous musical interaction. Therefore, it is reasonable to assume that if slow attack times allow a better perception of sound simultaneity this might also lead to a better ability to perform synchronous musical interaction

3 PILOT EXPERIMENT ON LATENCY TOLERANCE FOR DIFFERENT ATTACKS

Departing from the hypothesis that the PAT of individual notes can affect the quality of a Networked Music Performance, a pilot experiment was designed and conducted by the authors of this paper and Chris Chafe in December 2010 at the Center for Computer Research in Music and Acoustics (CCRMA). The main objective of this trial test was to investigate if better tolerance to communication latency can be achieved when individual notes are performed with slow attack instead of sharp/fast attack times.

The experimental setup consisted in two acoustic insulated studio rooms, in which two performers were subjected to a headphone based monitoring system as their acoustic communication loop.

The musicians performed on a Cello (musician A) and a Violin (musician B) and were asked to play the rhythmical structure presented on figure 2, which complies with previous experiments conducted at CCRMA [2] [4], for the sake of comparability.

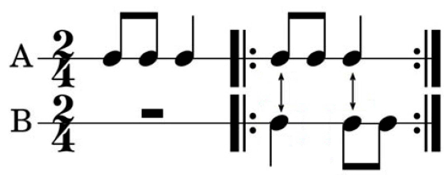

Figure 2: Rhythm used to test the latency performance with different attack times

When performed with a bow, both the Cello and the Violin allow the musicians to perform each note with a slow or a sharp/fast attack stroke. The experimental procedures consisted in getting the musicians to perform this rhythm with slow and fast attack times (as presented in figure 3), for different trials, in which latencies were artificially introduced in the feedback loop.

Since the musicians had to perform relatively long attack times, the adequate Tempo to avoid possible

---

[1] This particular experiment is entitled *"Effects of rate of onset on segregation"* and it is available, as part of the documentation for [12], from the website: http://webpages.mcgill.ca/staff/Group2/abregm1/web/downloadstoc.htm#21 [consulted in May 31, 2011]

overlaying of subsequent notes was set to 80 Beats per Minute (BPM).
Performing in such a low Tempo implies by it self a higher range of EPT (as suggested by the experimental research presented in section 1.1 of this paper) and therefore the latency values applied in each trial ranged from 25 to 110 ms.

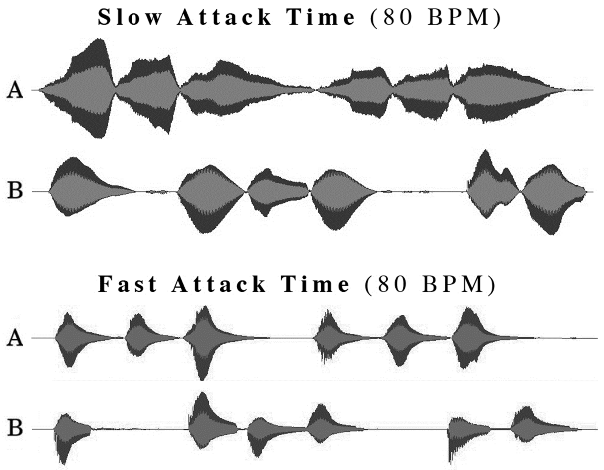

Figure 3: waveforms for performances with slow and fast attack times

For each trial, the initial procedure consisted in providing to booth musicians an electronically generated metronome beat set to 80 BPM, to which musician "A" would synchronize his initial performance followed by musician "B". Once both musicians concluded this initial synchronisation the metronome sound was muted and they would carry on with the performance for about one minute.
This procedure was repeated 28 times for slow and fast attack times, with latencies from 25 to 110 ms (with incremental steps of 10 ms).

4 RESULTS

The initial analysis of the audio files was performed individually for each musician, both for slow and fast attack time performances.
In order to have a perception of how each musician could maintain a steady rhythm, throughout the range of latencies introduced on the feedback loop, an analysis of the average BPM progression in time was executed with the Tempo and Beat Tracker analysis tool from *Sonic Visualizer*[2].
These results are presented in Figure 4 and 5, and it is clear that in booth cases the average Tempo is always slower than the initial 80 BPM metronome trigger, which is not really surprising since these average values convey the entire one minute performance, subjected to different latencies that disrupts the musicians steady Tempo performance.
However, these results also show that the average Tempo fluctuation for each trial decreases while latency increases (which is consistent with the inverse proportion between Tempo and EPT), but it is always higher for sharp/fast attack times than for slow attack times. In practice, this means that individually and in average, each performer is able to maintain a rhythm closer to the initial metronome Trigger (80 BPM) when performing sharp/fast attack bowed strokes.
Nevertheless, this analysis does not provide a straightforward perception of the performance quality, since it does not contemplate any information regarding the articulation between the two performances.

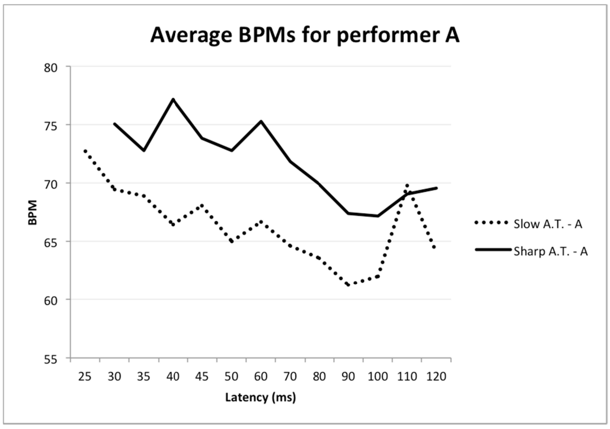

Figure 4: Individual Tempo average for performer A

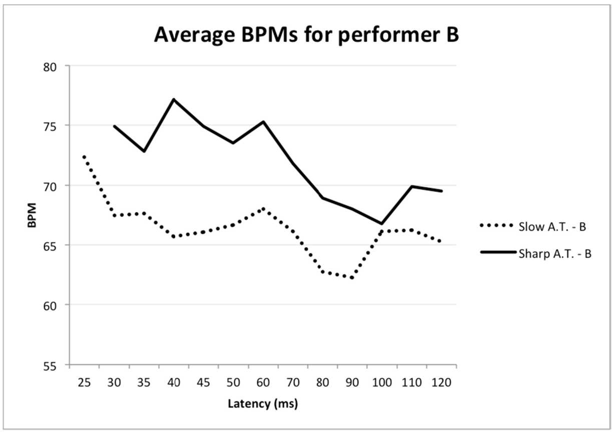

Figure 5: Individual Tempo average for performer B

In order to have an insight on the way each musician performed in relation to the other, a new analysis was performed using the software tool *MATCH: Music Alignment Tool Chest*, developed by Simon Dixon [14]. This software analyses the alignment of audio files using the on-line time warping (OLTW) algorithm [15],

[2] Sonic Visualizer software is developed at the Centre for Digital Music at Queen Mary, University of London: http://www.sonicvisualiser.org/ [consulted in May 31, 2011]

which was applied for the pairs of performances (slow and sharp/fast attack) in all different trials throughout the entire range of latencies.
The results of this analysis are presented in figure 6, and it is clear that the performance with slow attack is always closer to the value "1", which indicates a better proximity in the alignment of sound events from the two performances.

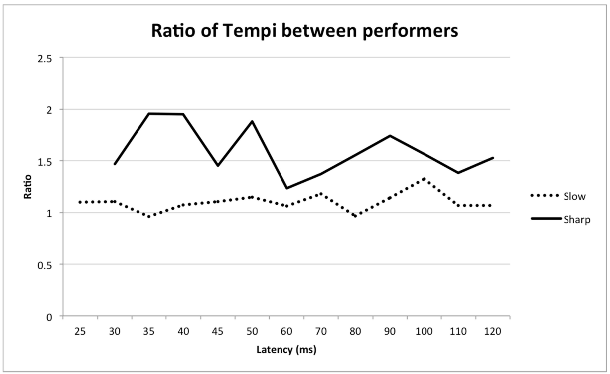

Figure 6: analysis of audio alignment using OLTW

This result clearly indicates that even with latencies as high as up to 110 ms, these performers can accomplish better outcomes in terms of rhythmical synchronisation with slow attack bowed strokes.

5 CONCLUSIONS

This paper presents recent research that introduces the dependency on perceptual audio features in the discussion of latency tolerance for Networked Music Performance. The relationship between latency tolerance, musical Tempo and dynamics has previously been demonstrated and briefly presented in the scope of this work. However, the main focus of this paper is concerned specifically to the relationship between tolerance to communication latency and the quality of a musical performance when individual notes are performed with slow attack instead of sharp/fast attack times. The results from the "Attack Time Pilot Experiment" suggest that for higher latencies than the generally established EPT (20-40 ms) [2], a better musical performance can be achieved when musical notes are played with slow attack times and slow Tempos. Nevertheless, this experiment is only a preliminary pilot test and further investigation remains to be done, specifically by expanding the experiment to a larger number of subjects and a larger scope of musical instruments. Furthermore, it remains to be determined what is the specific EPT range that can be achieved for this style of musical performance and if there is a functional relationship between the rise times of Attack in musical notes and the performance Tempo.

ACKNOWLEDGMENTS

The authors would like to thank "Fundação para a Ciência e Tecnologia - FCT" for the funding Álvaro Barbosa's Post-Doctoral Scholarship that allowed the realization of research published in his work, Chris Chafe and Hunter McCurry for their collaboration in the "Attack Time Pilot Experiment", Luis Gustavo Martins and Fabien Gouyon for their insightful comments on this work.

REFERENCES

[1] Holub, J., Kastner, M., Tomiska, O. 2007. *Delay effect on conversational quality in telecommunication networks: Do we mind?*. Wireless Telecommunications Symposium WTS 2007, Pomona, California-USA.

[2] Schuett, N. 2002. *The Effects of Latency on Ensemble Performance*. Undergraduate honors thesis - Stanford University.

[3] Chafe C., Gurevich, M. 2004. *Network time delay and ensemble accuracy: Effects of latency asymmetry*, Proceedings of the AES 117th Convention (San Francisco, CA: Audio Engineering Society).

[4] Lago, N and Kon, F. 2004. *The Quest for Low Latency*. In Proceedings of the International Computer Music Conference – ICMC 2004, Miami, USA.

[5] Bartlette, C., Headlam, D., Bocko, M., Velikic, G. 2006. *Effects of network latency on interactive musical performance*. Music Perception 24, 49 – 62.

[6] Barbosa, A. 2006. *Computer-Supported Cooperative Work for Music Applications*. PhD Thesis - Music technology Group; Pompeu Fabra University, Barcelona, Spain.

[7] Farner, S., Solvang, A., Sæbo, A., Svensson, U. P. 2009. *Ensemble hand-clapping experiments under the influence of delay and various acoustic environments*. Journal of the Audio Engineering Society 57, 1028 – 1041.

[8] Chafe, C., Cáceres, J-P., Gurevich, M. 2010. *Effect of temporal separation on synchronization in rhythmic performance*. Perception 39(7), 982 – 992.

[9] Barbosa, A.; Cardoso, J.; Geiger, G. 2005. *Network Latency Adaptive Tempo in the Public Sound Objects System* – Proceedings of the International Conference on New Interfaces for Musical Expression (NIME 2005).

[10] Stelkens, J. 2003. *PeerSynth: A P2P Multi-User Software with new techniques for integrating latency in*

real time collaboration. Proceedings of the International Computer Music Conference (ICMC2003).

[11] Spicer, M. 2004. *AALIVENET: An agent based distributed interactive composition environment.* Proceedings of the International Computer Music Conference (ICMC2004).

[12] Bregman, Albert S. 1990. *Auditory Scene Analysis: The Perceptual Organization of Sound*. 94. Cambridge, MA: MIT Press.

[13] Wright, M. 2008. *The Shape of an Instant: Measuring and Modeling Perceptual Attack Time with Probability Density Functions (If a Tree Falls in the Forest, When Did 57 People Hear it Make a Sound?)*. Ph.D. Dissertation, Stanford University, 19 – 20.

[14] Dixon, S., Widmer, G. 2005. *MATCH: A Music Alignment Tool* 6th International Conference on Music Information Retrieval, *Chest* – ISMIR 2005, London, England.

[15] Dixon, S. 2005. Live *Tracking of Musical Performances Using On-Line Time Warping*, Simon Dixon, 8th International Conference on Digital Audio Effects – DAFx'05, Madrid, Spain.