The Application of Flexilink in Multi-User Virtual Acoustic Environments

Calum Armstrong[1] and Jude Brereton[2]

[1]AudioLab, University of York, York, UK
e-mail: calum.armstrong@alumni.york.ac.uk
[2]AudioLab, University of York, York, UK
e-mail: jude.brereton@york.ac.uk

September 23rd 2016.

Abstract

In working towards a fully immersive multi-user Virtual Acoustic Environment latency between user interactions becomes a major concern. Allowing users to be physically separated whilst remaining together in the same virtual space requires the transfer of data between multiple environments to within a perceptually tolerable threshold. The "Virtual Singing Studio" (an ambisonic-based Virtual Acoustic Environment designed for singer rehearsal at AudioLab, York) is expanded to allow two performers to rehearse together in the same virtual environment whilst situated in separate rooms. Auditory data is sent from one user to another to facilitate the rendering of each performer's respective sounds in both Virtual Singing Studios. Due to the temporal unreliability of Audio-Over-IP networks, and further the need for minimal auditory delay between users, a new generation of Audio-Visual network communication is utilized - Flexilink. By prioritising Audio-Visual data packets over non-time critical data packets Flexilink helps to alleviate the perceptual strains of latency that are of particular importance.

1 Introduction

A Virtual Acoustic Environment (VAE) can be thought of as a computerised reconstruction (or auralization) of the acoustic properties of a space such that to a user within a VAE it sounds as if the auralized environment exists around them [1]. Key, however, is the implied interactivity of such a system. It is a system's ability to respond and adapt to a user's own input that best immerses the user inside the virtual reality.

A common approach is to simulate an acoustic environment by means of either measured [2, 3, 4, 5] or computer generated [6, 7] Room Impulse Responses (RIRs). An acoustically interactive VAE allows the user to sing or speak into the virtual space via a microphone input channel and hear the simulated response. Alternatively, or in addition to acoustic interaction, a system may implement realistic feedback to a user's physical movements, such as rotation or translation of the head and/or body. The result is that an acoustic source is localized with respect to the virtual space in which the user is placed, rather than to the user themselves i.e. as a user moves the virtual environment will appear to stay stationary.

Extending this idea to multiple users means simulating independent but virtually linked environments for each user. Not only should a user be able to move and hear their own voice in the virtual space, but they should also hear the voice(s) of the other user(s) auralized within the virtual environment. Importantly though, as this simulation is achieved in the digital realm the users do not have to physically be next to each other in real life to facilitate the effect of being in the same space.

Complications arise, however, in transferring data quickly enough between the hardware systems simulating each virtual environment. In order to simulate the voice of User A into the virtual environment of User B effectively, a live input must be sent instantaneously from User A's microphone to the simulator of User B. As the physical distance between these two nodes increases this task becomes harder to accomplish.

2 Motivation

The Virtual Singing Studio [4, 5] was originally devised by Brereton to provide controllable virtual room acoustic environments in order to help investigate performance variations of singers in varied spaces. Instead of transporting singers to different physical locations the VAE facilitated almost instantaneous virtual transportation between environments simulated from measured RIRs of the venues tested.

In practise, the concept lends itself easily to performance rehearsal. The ability to simulate a venue for a musician may soon mean that a singer could realistically rehearse for an upcoming show in the Sydney Opera House whilst still based in New York. Often, however, performers do not perform alone. A well implemented Multi-User VAE should allow multiple performers to rehearse together in the same virtual space whilst, potentially being located in different countries across the globe.

Furthermore, technologies such as conference calling would benefit from the generation of a realistic virtual space within which voices are placed and localized at realistic positions, rather than simply being output from one/two arbitrarily placed loudspeaker(s).

3 Set-Up

3.1 Virtual Singing Studios

Two VAE rigs were constructed, each simulating an environment for a single user. The original Virtual Singing Studio was utilized alongside a second simplified set-up. Whilst the original studio comprised a 16-point octagon-cube ambisonic speaker rig the partnered studio comprised only an 8-point horizontal octagon array. The two Virtual Singing Studios were constructed in separate but adjoining rooms.

Each singer wore a cheek mounted microphone and was able to hear not only their own voice, but also the voice of the other user, simulated within the virtual environment. Parameters for these simulations were set in the accompanying independent software that fed audio to each studio.

3.2 Aubergine Network

Studios were linked via two Aubergine network boxes [8] connected by a hardline Ethernet cable. The Aubergine boxes interfaced AES audio connections with a Flexilink [9] network connection. A Flexilink network differentiates data-packets based on their need for time-critical transmission. In a standard network, data-packets are sent in the order they arrive, meaning the transmission of any other data, say the downloading of a text file, may interrupt an audio stream causing glitching / latency. Flexilink, on the other hand, initiates semi-permanent links between critical device inputs / outputs ensuring audio / visual data streams remain uninterrupted, juggling the transmission of non-time critical data in the leftover bandwidth.

Two-way audio / visual communication was set up via the Flexilink network. The microphone input of each user was both sent to and received from the other.

3.3 MAX MSP

Each VAE was generated by an independent copy of an almost identical piece of software running on MAX, which provided a quick and easy facility to alter / adapt the program code to meet the needs of the application. The programs differed only in their processes for decoding the ambisonic audio generated within the programs for different speaker array layouts. The VAE MAX programs ran on two separate computers and were interfaced to the speaker rigs / Aubergine network boxes via multi input / output fireface audio interfaces [10].

The programs accepted an audio input for the primary user of each VAE (taken directly from the head mounted DPA 4066 microphone) as well as inputs for partner users (received via the Aubergine network box links). Each mono audio stream was then separately convolved with pre-chosen directional B-format RIRs which “placed” the sound sources within the virtual space. Convolution outputs were subsequently summed and decoded for the speaker array of the relevant loudspeaker distribution.

4 RIR capture

Convolving with correctly measured RIRs is essential in simulating a realistic 3-dimensional Virtual Acoustic Environment. For each user voice simulated within each virtual environment an RIR is required which represents the source-to-listener position being simulated. As a singing duet was simulated in these tests two singing positions taken from within a vocal quartet layout were chosen for the basis of measurements.

B-format RIRs were captured at each potential source position (1-4), with the source placed at mouth height and receivers placed at ear height at all possible receiver positions (see example recording layout

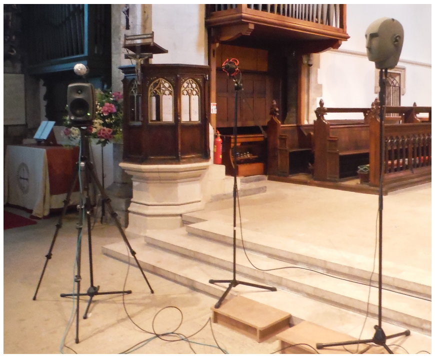

Figure 1: RIR recording set-up showing various reciever positions being recorded simultaniously for a single source position (left)

		Primary User	
		User A	User B
User Input	User A	2 → 2	2 → 3
	User B	3 → 2	3 → 3

Table 1: RIRs, shown in terms of their respective source - receiver positions, convolved with each user input to simulate a coherent virtual space for each Primary User.

in Fig. 1). From these recordings 16 RIRs were generated representing the RIRs for a singer listening to their own voice, and a singer listening to the voices of the three other members of the quartet respectively.

Within the Virtual Singing Studio software it was possible to select corresponding RIRs for convolution with user inputs in order to simulate a coherent virtual acoustic space. The set-up tested here simulated two users standing side by side, facing forward, each hearing the other user's voice coming from either their right or left respectively. Table 1 shows the RIRs chosen for convolution with user inputs for each of the primary user's programs in terms of their source-receiver positions for the tests carried out.

5 User Tests

Initially, informal tests were undertaken to determine the latency of the system. Measurements were taken of the time difference between an impulse received at a user's microphone and the same impulse sound arriving in the partnered system's loudspeaker response. The results are shown in Table. 2.

Set-up	**A → B**	**B → A**
Mic → Speaker	40ms	39ms
Mic → Center	43ms	44ms
Travelling Time	3.5ms	4.4ms

Table 2: Latency times measured between an audtory user input and its reception at a partnered VAE. Measurements are made at both the receiving studio's speakers, and in the center of the speaker array. The time taken for a sound to reach a user after being emmitted from any speaker is given by 'Travelling Time'.

Minor differences in latency time between systems can be attributed to computer processing power (System A had 8GB of RAM, system B had 4GB of RAM) and speaker array radius (system A had a radius of 1.5m, system B had a radius of 1.2m).

Perceptual user tests were also carried out using a methodology developed for a previous similar study [11]. Users were asked to perform a short piece of a cappella music, both in unison and in cannon (taking it in turn to lead the cannon). This short piece was repeated twice, whilst users a) stood together in real life and then b) were physically separated in real life but virtually together via the VAE. Users were tasked with maintaining a steady tempo whilst keeping in time with each other. A selection of results are shown in Fig. 2.

6 Test Results

The tempo maintained by each user throughout each recital was calculated via plosive inter-onset interval analysis and plotted for each note sung over time. Lines of best fit were approximated to show the overall variation in tempo for each user over each performance.

Despite significant variations being seen on a note-by-note basis results show in general a good consistency of tempo over time for both real and virtual environments. If system latency had presented a major issue for the singers it would have been expected that the performance tempo would have gradually fallen over time systematically throughout the virtual set-ups when compared to the real life set-ups. Despite a clear overall drop in tempo in 'VSS U1 Tempo' this is not seen in either of the other two virtual tests and can in part be explained by the high first few anomalous results.

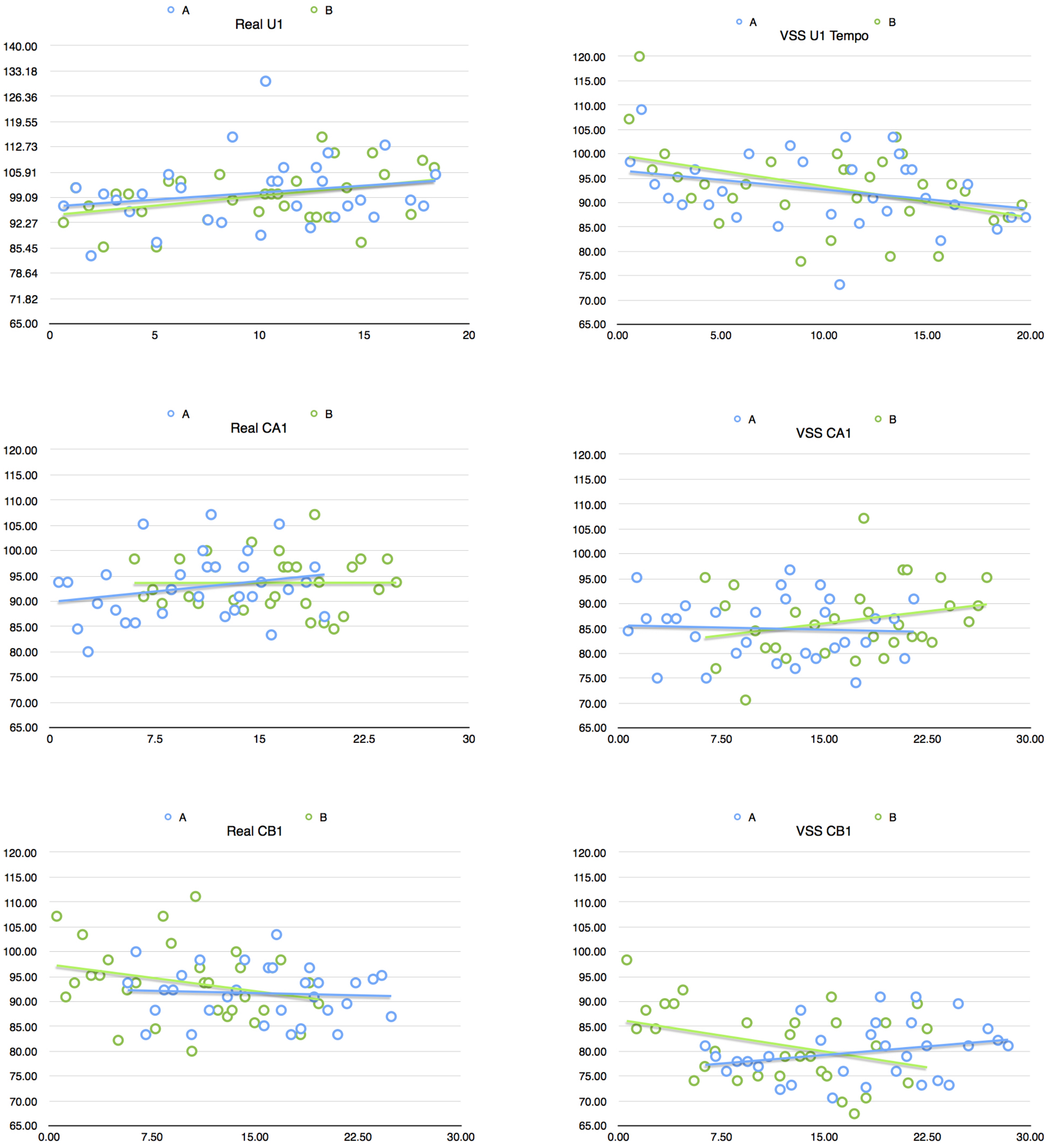

Figure 2: A selection of extracted tempo over time results for each of two subjects singing in: unison (U) (Top Row); cannon, Subject A leads (CA) (Second Row); cannon, Subject B leads (CB) (Third Row). Tests were repeated with subjects performing together in real life (Real) (First Column) and virtually performing together via a Multi-User VAE (VSS) (Second Column). Tempo is shown on the vertical axes in Beats Per Minute (BPM) whilst time is shown on the horizontal axes in seconds. Automatically generated lines of best fit are shown for each subject for each test.

7 Future Work

Ideally, integration of this two-node Flexilink network to a larger communication network would have been made. However, in order to handle the audio / visual data as efficiently as possible each and every node the data streams passed through had to implement Flexilink. Sacrificing this constraint would have allowed the data streams to fall victim to standard IP latencies at at least some of the network's junctions. As a larger Flexilink network was not available, this was unfortunately unable to be extensively tested.

8 Conclusion

Although the results obtained here are by no means categorical, they do imply that such a network connection as implemented by the Aubergine network boxes and Flexilink protocol provide a method of communication sufficient for multi-user VAEs. Further latency improvements should be considered with respect to the VAE generation software as well as data transmission. Nevertheless, Flexilink offers significant improvements over previous generations of network audio communication, as presented by Geering [11]. Further work will be carried out into the application of such technology over a wider network and the streamlining of such an approach.

9 Acknowledgements

Throughout this work John Grant from Nine Tiles [12] played an essential role with regards to both the the supply and operation of the Aubergine network boxes. Without his support such a project would not have been possible. Thank you.

References

[1] C. Armstrong, "Towards a technically accurate soundfield based interactive virtual acoustic environment," Literature Review, University of York, 2015.

[2] M. Rokutanda, T. Kanamori, K. Ueno, and H. Tachibana, "A sound field simulation system for the study of ensemble performance on a concert hall stage," *Acoustical Science and Technology*, vol. 25, no. 5, pp. 373–378, Mar.2 2004.

[3] W. Woszczyk, D. Ko, and B. Leonard, "Virtual acoustics at the service of music performance and recording," *Archives of Acoustics*, vol. 37, no. 1, pp. 109–113, May 2012.

[4] J. Brereton, D. Murphy, and D. Howard, "A loudspeaker-based room acoustics simulation for real-time musical performance," in *25th Audio Engineering Society UK Conference 2012*, vol. 1. Audio Engineering Society (AES), Mar.25-27 2012, paper 09.

[5] ——, "The virtual singing studio: A loudspeaker-based room acoustics simulation for real-time musical performance," in *Baltic-Nordic Acoustics Meeting 2012*. SINTEF, Jun.18-20 2012.

[6] M. Kleiner, B.-I. Dalenbäck, and P. Svensson, "Auralization-an overview," in *91st AES Convention (preprint 3119)*. Audio Engineering Society (AES), Oct.4-8 1991.

[7] T. Lentz, D. Schröder, M. Vorländer, and I. Assenmacher, "Virtual reality system with integrated sound field simulation and reproduction," *EURASIP Journal on Advances in Signal Processing 2007*, vol. 2007, Jan.3 2007.

[8] Aubergine. Product Website. Nine Tiles. [Online]. Available: http://www.ninetiles.com/Aubergine.html [Accessed: Sep. 03 2016]

[9] Flexilink. Product Website. Nine Tiles. [Online]. Available: http://www.ninetiles.com/FlexilinkIntro.html [Accessed: Sep. 03 2016]

[10] Fireface UFX. Product Website. RME. [Online]. Available: http://www.rme-audio.de/en/products/fireface_ufx.php [Accessed: May 06 2016]

[11] S. Geering, "The virtual singing studio: Expansion to multiple users," Master's Thesis, University of York, 2015.

[12] Nine Tiles. Company Website. Nine Tiles. [Online]. Available: http://www.ninetiles.com.html [Accessed: Sep. 09 2016]