ACOUSTICS OF VIRTUALLY COUPLED PERFORMANCE SPACES

Braxton Boren

American University
Audio Technology Program
4400 Massachusetts Avenue
Washington, DC 20016
boren@american.edu

Andrea Genovese

New York University
Music and Audio Research Laboratory (MARL)
35 W. 4th St.
New York, NY 10012
genovese@nyu.edu

ABSTRACT

Many different musical applications, including remote sonification, sound installation, augmented reality, and distributed/telematic music performance, make use of high speed Internet connections between different performance spaces. Most of the technical literature on this subject focuses on system latency, but there are also significant contributions from the acoustics of all rooms connected: specifically, smaller auxiliary rooms will tend to introduce spectral coloration, and the "main" larger volume will send more reverberation to the off-site performers. Measurements taken in two linked networked sites used in telematic performance show that both of these issues are present. Some improvements are suggested, including physical room alterations and equalization methods using signal processing.

1. INTRODUCTION

Many different musical applications, including remote sonification [1], auditory display [2], augmented reality [3], and distributed/telematic music performance [4, 5], make use of high speed Internet connections between different performance spaces. These share the goals of fostering greater scientific, personal, and musical interaction over distance by connecting scientists, educators, musicians, and audience members who would not normally be able to communicate in real time. These distributed sound interactions share the goal that such virtual connections should sound as perceptually naturally as possible [6], ideally indistinguishable from a comparable interaction within a single acoustic space, wherein two performers are only distributed across a few meters of a natural stage [7].

In all these areas, which we will group together under the umbrella term Networked Audio Platforms [8], the chief barrier to achieving 'naturalness' has been latency in the time required for a signal to travel from one participant to another, and as such it is also the chief problem addressed in the corresponding literature [9, 10, 11]. Bartlette and Bocko specify this as "external latency" as opposed to "internal" latency or timing variance due to differences between individual performers [9]. This external latency can be further subdivided into transmission delay ("propagation latency") and any additional delay introduced by signal-

processing demands during transit ("system latency") [7]. Finally, propagation latency may be either acoustic or electronic, depending on whether the signal is traveling between a sound source and a transducer, or between two transducers via the Internet.

In the past two decades, increases in computational power have reduced system latency for simple transmissions, making propagation latency the controlling variable for final system latency. Depending on the exact distances in question, common electronic propagation latencies are often on the order of tens of milliseconds [7]. Some have made the case that because there are non-networked (i.e. acoustic) sound environments in which such latencies also occur, that electronic path latencies of similar magnitude are satisfactory [9]. This reasoning requires that all acoustic sound sources are close-miked with no appreciable sound path between source and microphone, and likewise that all listeners are wearing headphones, with virtually no time delay between the transducer and the listener' ear drums. In practice, both of these assumptions are untrue in various networked audio settings, meaning that the total propagation latency may be larger. In addition, it is often assumed that electronic signals travel at the speed of light, but this is also not practically achievable due to certain Internet infrastructure limitations [12]. Acceptable latency values for musical applications are variously reported as up to 30 ms [13], up to 50 ms [14], or higher [9], depending on how the criterion for acceptability is defined.

Despite this understandable historical focus on electronic propagation and system latency, Network Audio Platforms are also profoundly affected by the room acoustics of the networked spaces at either end (or more than two ends in many cases). In many cases the physical acoustics of the rooms at either end of the transmission line may affect the perceptibility of the delays due to transmission. Though individual acoustic effects have been addressed at times in the literature, there exists no single systematic treatment of the relationship between the acoustic characteristics of virtually linked spaces and the final audio system that results from their combination. This paper provides a background on the acoustic effects present in networked spaces, as well as the psychophysics of coupled acoustic volumes and musicians' responses to different types of non-networked acoustic spaces. It will then present a unified framework for addressing systematic musical asymmetries that may arise from differences in the acoustic responses of networked performance rooms.

2. BACKGROUND

Though the literature on networked audio platforms focuses primarily on propagation latency, there are several references to room acoustics, albeit mostly in the context of addressing latency. Many researchers have noticed that at the present, it is impossible to remove all latency issues from transmission, so they propose a compromise between the real world and the effects of the transmission line. This 'compromise' strategy includes "using networked time delay in a musical fashion rather than constantly trying to counter it" and "combining the acoustics of physical spaces with the acoustics brought by the network itself" [10]. Chafe [12] designed a system to use feedback delays within a networked audio platform to trigger comb filters, which were later processed with an all-pass delay filter to create a single virtual reverberation on both ends of the network, thus using a real acoustic phenomenon to mask an audio problem. Gang et al. [11] showed that reverberation can mask some deleterious effects of transmission latency. While Braasch [7] does caution that it is important for networked performers to have a greater awareness of real-world sonic environments, he also stresses that "unique affordances exist for telematic music systems, which is a basic requirement for them to serve as a platform for new types of music."

The label of 'compromise' above is not meant as a pejorative - indeed this approach is responsible for a great improvement in the smoothness and fluidity of networked audio performances in the past two decades. However, it is not advisable to let the necessary limitations of the present obscure our vision for the future: to take an example from another area of audio technology, for years headphone design focused on free-field or diffuse-field equalized headphones because these were intended to reproduce stereo audio mixed for loudspeakers [15]. But as the demand for high-quality binaural audio for virtual reality has increased, it is clear that more comprehensive inverse filtering will be necessary to remove the response of headphones entirely to provide a perceptually blank slate for dynamic binaural synthesis [16]. In the same way, a 'compromise' networked audio brings participants on either end into a virtual, neutral acoustic space which is not necessarily similar to the physical rooms on either end of the transmission. But a future system might offer either a sort of "acoustic gateway" between two sites, wherein one participant receives the same 3D sound field as a listener in physical room miles away, or an "acoustic holodeck" [17] where participants meet in a neutral virtual space, but one that is fully changeable and not dependent on the acoustic or audio characteristics of the physical performance spaces or their distance apart. Neither of these goals are attainable at present, and they are constrained not just by propagation latency but also by room acoustics. To better understand the impediments to improved networked audio platforms, it is important to understand the acoustic constraints imposed by the physical coupling of two separate volumes.

3. ACOUSTIC COUPLING EFFECTS

If a signal $x(n)$ is produced in a physical room and recorded at a microphone in the same room, $R1$, the resultant recorded signal $y(n)$ will be equal to the convolution

$$y(n) = x(n) * h_{R1}(n), \qquad (1)$$

where $h_{R1}(n)$ is the room impulse response (RIR) representing the acoustic system from the sound source, through the room, and

into the microphone (thus accounting for any filtering characteristics of the microphone itself). The input sound $x(n)$ will experience the room decay present in the RIR $h_{R1}(n)$. Assuming the room is reasonably diffuse, the RIR will exhibit an exponential decay, which when plotted in dB will appear as a straight line. The slope of this line determines the reverberation time, T_{60}, of the room.

If we imagine a second adjacent room $R2$ with corresponding RIR h_{R2} and a longer reverberation time, and then create an opening or aperture of surface area S between $R1$ and $R2$, then this 'open-window' aperture appears, from the perspective of a single room, as an additional source of absorption such that each room's total absorption values becomes

$$A_{iS} = A_i \alpha_i + S, \qquad (2)$$

where A_i is the amount of surface area in the ith room, and α_i is the average absorption coefficient of the ith room [18].

However, the sound traveling through the aperture is not lost completely, as with an open window - instead it enters the second room, whose sound will likewise spill back into the first room. The phenomena described here is called acoustic coupling, and the system consisting of two coupled volumes will exhibit a decay curve that is a combination of the decay curves of both constituent volumes. If we now measure the impulse response from a source to a receiver, both in $R1$, with $R1$ partially coupled to a more reverberant $R2$ through an aperture, then the resulting impulse response will have a non-linear decay curve, made up first of the steeper decay rate of h_{R1}, while the later part of the curve will represent the more gradual decay rate of h_{R2}. As the curve transitions between these linear portions, there will be a bend which has a greater dB value than either of the individual rooms' decay slopes, as shown in Fig. 1. This is known as the double-slope effect (DSE) or a double-slope decay (DSD) [19, 20].
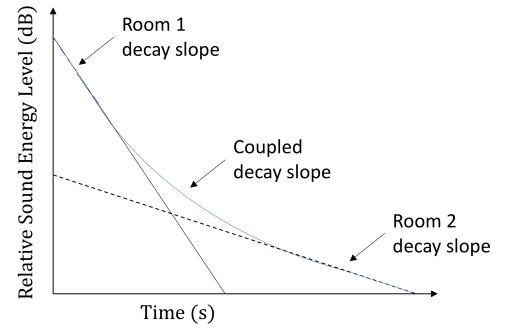

Figure 1: Double-Slope Effect resulting from the coupling of two rooms with different decay rates.

The early and late portions of the decay are respectively given by

$$L_{early}(t) = -\left(\frac{60}{T_1}\right)(t) \qquad (3)$$

and

$$L_{late}(t) = -\left(\frac{60}{T_2}\right)t - 10\log_{10}\left(\frac{A_{1S}A_{2S}}{S^2}\right), \qquad (4)$$

where T_i is the independent reverberation time of the ith room, and t is the time index [18]. Note that the second term in Equation 4 implies that impulse response of the coupled system is not equivalent to the convolution of the impulse responses of each room alone.

Coupling effects need not be restricted to only two connected rooms, as more adjacent volumes and apertures could produce a non-linear decay made up of three or more piecewise linear components, each corresponding to the decay rate of one of the individual rooms [18]. Because of the nature of the DSE, it is necessarily more complicated than a single room's decay and cannot be described by a single number as can a diffuse volume using T_{60}. Different parameters to measure the degree of coupling have been suggested, but all vary in effectiveness depending on the ratio of the linear slopes and the transition zone between them [20].

In the example above, if we instead put the source and receiver both in $R2$, the more reverberant of the two rooms, its (more gradual) decay rate will also serve as the initial decay rate, but because the sound will decay much more quickly in $R1$, there will be no DSE from the perspective of $R2$. This introduces an interesting asymmetry that will be important for later analysis: *dry rooms can hear reverberant rooms coupled to them, but reverberant rooms cannot hear dry rooms coupled to them* [18]. Reverberant volumes can act as a sort of time capsule, keeping sound around past when it would normally have decayed in the drier room, and letting that sound back into the dry room. For this reason, coupled volumes have been studied in the context of room acoustics, especially regarding dry spaces (like balconies or small chapels) connected to more reverberant spaces (such as concert halls or the nave of a cathedral) [21, 22]. However, the concept of coupling can easily be extended to two spaces which are not acoustically joined, but rather connected via networked audio.

4. COUPLING IN NETWORKED AUDIO PLATFORMS

The situation for Networked Audio Platforms has many distinct differences from the purely acoustic coupled volumes presented so far, but there are also many similarities. At the most basic level, there are still at least two different rooms which are hearing sound from each other, and these rooms likely have somewhat different acoustic properties. The range of different combinations is too large to address every possibility, so instead we restrict ourselves to a fairly simple setup which is used in many telematic/distributed performances. In this scenario, we assume there is a single 'main' volume containing multiple performers and an audience, and a second 'off-site' volume connected via networked audio, containing a single performer in another location who is joining in with the performance virtually. Just as before, the combined coupled system's impulse response will not be identical to the convolutions of the individual impulse responses, based on differences introduced not through an aperture but through the audio system connecting them.[1]

Before considering the acoustic issues introduced, we must define some basic features of the two rooms. Again, individual

situations may differ, but because of the presence of multiple performers plus an audience in the main site, we may usually assume this location is the larger of the two volumes. The off-site location need not be very big, and may be assumed to be roughly the size of a practice room or iso booth. Similarly, the off-site performer needs only an instrument, a computer, a microphone, and headphones (which are not strictly necessary but will be assumed for this example), while the performers at the main site, along with the audience, are receiving audio over loudspeakers.

4.1. Frequency Domain Effects

What acoustic issues will arise in such a system, irrespective of the propagation delay between sites? First let us consider the frequency domain: if the smaller room is a well-designed iso booth built to favorable dimensions and with no parallel surfaces, it may have a reasonably smooth frequency response due to the wide distribution of resonant modes through the audible spectrum. If instead it is a simple rectangular volume with flat plaster walls, it may exhibit significant modal resonances which will distort the timbre of the signal being sent to the main room.

Conversely, the larger main volume will have a low Schroeder frequency, meaning that most if not all the reflections within the room can be considered diffuse and evaluated chiefly in the time domain rather than by considering individual room modes [23]. This means there are likely to be no large resonant peaks or notches in the main room's frequency response, either from the perspective of the audience or the off-site performer. However, the sound coming from off-site may be significantly affected, depending on the geometry of the smaller room. These frequency-dependent effects, however, can be easily compensated for via regularized inversion in the frequency domain, a technique that is widely used to equalize headphones, loudspeakers, and rooms [24, 25]. Even though the larger volume is likely to have a smoother spectrum, the reverse may also be applied to equalize the main performers' audio before reaching the off-site room.

This ability to get near-perfect removal of frequency effects means that the off-site performer may also be removed from their own room and put into the main volume, at least as far as the frequency domain is concerned. Since the off-site performer is listening through headphones, they will hear both their own instrument filtered through the off-site room, as well as their delayed version which has been equalized. If EQ is being applied in both directions, theoretically this delayed version could be free of coloration due to either of the rooms, within the constraints determined by regularization. One caveat is that such equalization assumes that the frequency response of both rooms has been measured in advance, which is not always the case.

4.2. Time Domain Effects

In the time domain, the situation is more complex: the smaller site will usually have a very small T_{60} value, unless it is extremely reflective. The larger main site will likely have a longer reverberation time, especially if it is a site intended for live music performance. Because of that, there are various sound paths that need to be considered: one-way paths, from one site to the other, and round-trip[2] paths, from one site to the other, filtered through the second site,

---

[1]Note that in the case of virtual coupling, the amount of additional absorption from the first room's perspective due to an 'aperture' is insignificant, since the 'aperture' is just a microphone with little acoustic absorption. However the spillover effects from room to room proceed as with physical coupling, increased to some extent because of the amplification of the virtual signal received at the microphone.

[2]In some distributed performances [26] audio is designed to flow only one-way and the return path is disabled, in which case the round-trip analysis will not apply.

and back to the sound's origin. These different paths are shown in Table 1. In addition, for reference we will sometimes consider the single-room impulse responses for either space alone.

Table 1: Sound paths

One-Way	Round-Trip
Off-Site to Main Room	Off-Site Back to Off-Site
Main Room to Off-Site	Main Room Back to Main Room

In general, the discussion of these four main paths involves the amount of reflection and reverberation added by sound traveling through both rooms. This merits the question of whether blind dereverberation may be applied in the time domain as easily as regularized inversion in the frequency domain. In principle dereverberation may be achieved similarly to frequency equalization, but since RIRs are orders of magnitude longer than their frequency domain equivalent, dereverberation is much more computationally complex. Since different source positions determine slightly different RIRs, many dereverberation techniques focus on a blind determination of the best inverse filter for the total audio signal received, a problem which is still unsolved. Exact inversion may actually degrade the input signal, whereas algorithmic processing adds significantly to system latency [27]. In addition, many existing techniques focus only on improving signal-to-noise ratio for speech intelligibility, which is applicable to some (e.g. teleconferences) but not all (e.g. distributed music performances) Networked Audio Platforms [28, 29]. While multiple input channels of audio can improve the output filter [30, 31], this also increases computational load. In the future, multiple audio channels may be desirable both for increased dereverberation and for improved immersion for performers, whose preferences include spatial as well as temporal placement of sound reflections [32]. While dereverberation certainly poses an important area of future exploration, at present it is not possible to cheaply and quickly remove the reverberation from a signal sent from one room to another, implying that these rooms are in a sense coupled together similarly to adjacent acoustic volumes connected by an aperture. Keeping this in mind, we will apply the theoretical framework developed in Section 3 to the virtual coupling in our imagined musical performance.

4.2.1. Off-Site to Main Room

Since the signal from the off-site performer receives a much lower T_{60} value than the main room, it emerges from a loudspeaker in the main room dry, and then experiences the longer reverberation of the larger volume. In this sense it is similar to the instruments of the performers in the main room, whose signal emerges dry – either from acoustic instruments or from loudspeaker-based virtual instruments – and is convolved with the RIR of the main volume. While some early reflections will be mixed with the off-site performer's direct sound, these will be lost in the larger mix of reflections in the main volume and will not be perceptible to the audience.

4.2.2. Main Room to Off-Site

The one-way path for sound originating in the main room and traveling off-site involves a larger reverberation time in the larger volume, which will then be heard in the signal received off-site. Assuming (as we are) that dereverberation is not a practical option,

this creates an asymmetry in the one-way sound paths: the off-site performer hears the added decay of the main room in addition to their own, but those in the main room do not hear the time decay of the off-site room.

4.2.3. Off-Site Back to Off-Site

From the perspective of the off-site room, the performer's own sound is transmitted through the smaller room, through the network, into the main room, which then acquires the reverberation of the larger space before being fed back into the network and reaching the off-site performer. This is a feedback echo, which is used in the 'compromise' strategy described above to create a single, shared reverberation time between both spaces [12]. If the virtualized T_{60} is greater than that of the main room, then the virtual reverberation will indeed be heard equally by performers on both ends. If, however, the main room's T_{60} is greater, then this will form a DSD similar to Fig. 1. The early decay time (EDT) may still be perceived as that of the shorter decay rate [18], but the late field reverberance will likely be perceived as the longer T_{60} value of the main volume, depending on the ratio between the two decay rates. This suggests that the minimum T_{60} of the virtually coupled system is constrained by the greatest T_{60} of any of the rooms being joined, i.e. $T_{system} \geq T_{main}$ in our example. In addition, this affects the stable range of achievable decay rates, constraining the types and tempos of music that may be realized in a reverberant acoustic system.

4.2.4. Main Room Back to Main Room

From the perspective of the main room, the performers' signal acquires some reverberation in the larger volume before being sent off-site. Since the off-site performer is wearing headphones, this signal will not be fed back, and thus there is no significant round-trip path in this case. However, the performers will still hear the T_{60} of the main room acoustically (more strongly than the off-site performer). If a virtualized reverberation is imposed on the network, then it will not be heard separately unless its T_{60} is greater than that of the main volume, which will also cause a DSD.^{3} Thus the acoustics of the larger volume become a controlling factor for the performance of both the on-site and off-site musicians.

5. MEASUREMENT RESULTS

In order to explore and validate the theoretical foundation under the given assumptions, the case scenario illustrated in section 4 was measured in the time and frequency domains. Two candidate locations known for having been previously used for distributed music performances were chosen within the NYU Steinhardt Education building. The Frederick Loewe Theatre (294 seats, dimensions approximately 17x21x6 m, volume approximately 2180 m^3) was selected as the main site, while the Research Laboratory on the 6th floor of the same building (rectangular, hemi-anechoic, dimensions 4.5x3.7x2.5 m) was selected as the smaller off-site location for remote performance.

In both rooms, a Genelec speaker was placed one meter away from a C617 Josephson omnidirectional microphone. Three 2-second exponential sine-sweep measurement signals (ranging

---

^{3}It is also possible that a performer very close to a loudspeaker could perceive the virtualized EDT, but we are principally addressing the room as a whole.

from 20Hz to 20kHz) as described in [33] were controlled from a central studio location able to patch to both environments. The sine-sweeps were first measured independently in both rooms, and then through "one-way paths" in both directions, meaning that a sine-sweep reproduced in the main room was captured in the main room, routed to the off-site room and measured, and vice versa. Finally, round-trip paths were measured from a source in one room, through the other room, and back to the first - this was done by playing the "one-way" sweep recording from the opposite room's speaker and recording its response in the originating room.

5.1. Frequency domain

Despite the claim that inverse filtering of room effects could remove frequency coloration from off-site volumes, in practice real-time filters to implement this equalization will add somewhat to the propagation latency between the two sites, and are therefore often not used in telematic performance. However, the measurements in fig. 2 illustrate the frequency responses of the two measured rooms, plus that of a source in the off-site room and measured at a microphone in the main room. What comes primarily to attention are the strong notches captured within the off-site room, probably caused by the resonant modes typical of rectangular spaces. The same notches are also captured in the sound transmitted from the off-site room to the main room. This suggests that the incoming transmissions to the main room present timbre coloration introduced by the acoustical character of the off-site room. In contrast, the response of the main room alone indicates a more diffuse field and less dominance of any particular resonances.
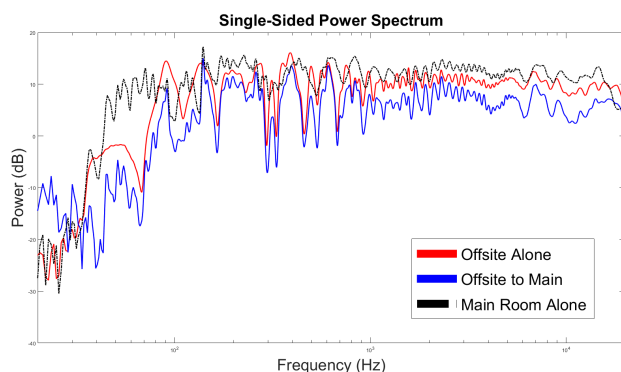

Figure 2: Frequency spectrum resulting from the coupling of a small offsite space with a larger main room, smoothed according to [34]. The red line represents the off-site room's transfer function alone, the blue line is the spectrum of a source in the off-site room measured from a microphone in the main room, while the dashed black line represents the spectrum of the main room alone.

Based on these significant notches of more than 10 dB, some equalization seems advisable for use in musical performance. A simplified inverse IIR filter or short FIR filter could be used to correct for the biggest spectral coloration without large increases in system latency. In addition, some networked audio applications using spatial audio may already require real-time convolution for implementing binaural filters on transmitted audio – in this case, the addition of a single static inverse filter does not require extra FFT operations, but only an additional vector multiplication in the frequency domain, which is not a large computational load

increase. In addition the convolution of a binaural filter with the static room inverse filter could also be pre-computed, which would further reduce latency at the expense of additional system memory.

5.2. Time Domain

The backwards-integrated Schroeder decay curves were calculated by octave band for the off-site room alone, the off-site room to the main room, and the round-trip path from the off-site room back to itself. The results for the octave band centered on 1000 Hz are shown in fig. 3. The results show an expected agreement with the Double-Slope Effect: as expected the T_{60} value for the small off-site room is approximately 0.1 s. At the receiving end, the sound is played back through a loudspeaker and mixes with the hall's reflections. Given the slower decay rate of the concert hall, both the one-way and round-trip RIRs show a distinct bend indicative of coupling between the faster decay of the off-site room and the slower decay of the main volume. In the late field, the captured sound presents now approximately the same decay characteristics as the one reproduced in the main room. The off-site performer will thus be subjected to both the resonant modes of the smaller space, and the reverberation effects of the larger hall.
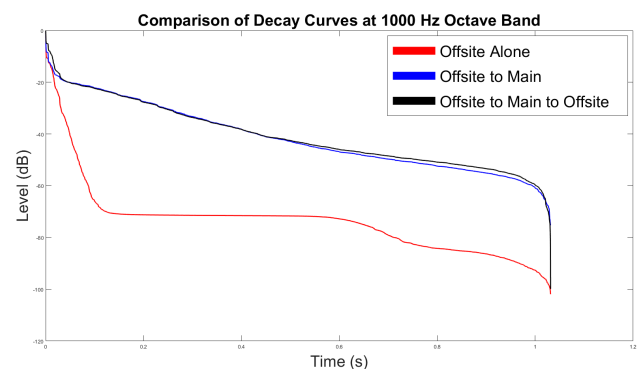

Figure 3: Decay curves measured for coupled rooms. The red shows the decay in the off-site room alone, the blue curve is the signal decay from a source in the off-site room measured at a microphone in the main room, and the black curve the resulting round-trip loop from a source in the off-site room, through the main volume, and measured back in the off-site room.

6. CONCLUSION

Due to the nature of acoustic coupling and DSD, rooms which are virtually linked will also transfer their acoustic parameters to each other to a certain extent. In the frequency domain, it is relatively simple to remove the worst modal resonances of small spaces' audio signals. But in the time domain, it is currently not possible to adequately remove reverberation from a larger space's signal, and applying existing dereverberation techniques adds to system latency considerably. This poses significant tempo constraints on the performers, who will tend to play slower in the presence of greater reverberation [35, 36]. Networked audio often tends to feature slow tempos, and reverberation is thought to help mask feedback delay, so this is not always a problem. However, it puts a hard constraint on the overall tempo space that may be achieved via telematic performance. To use a historical analogy, it was not

a problem to perform slow monophonic chants in highly reverberant churches, but the development of faster, more complex music necessitated drier spaces with greater acoustic clarity [37].

On top of the very real constraints imposed on distributed music performance by propagation delay and feedback echoes, the acoustic properties of any large space used during the performance are a permanent constraint on the composite Networked Audio Platform. This can of course be addressed by adding sufficient absorption to any large volumes to create extremely low T_{60} values. Though performing in a very dry space can also cause performers to lower their tempo [38], additional virtual reverberation can always be added in.

This suggests a problem, in that many distributed performances' main (i.e. audience-containing) room is often intended for acoustic performance and thus has a T_{60} between 1-2 seconds at mid-frequency. One possibility for distributed performance is to simply remove the large room, sending out an Internet live stream to an audience that is equally distributed, free to listen wherever they are. However, this also removes any communal aspect to the experience of the music, and would diminish the experience of attending a performance rather than simply consuming it like any other Internet video. A better solution is to always place the audience for a Networked Audio performance in a dry space, such as those intended for rock music or electroacoustic performance. This removes the need for dereverberation while ensuring that, whatever virtual acoustic is painted on the audio transmissions, all performers are starting from a blank slate.

Aside from the discussed case scenario, a further question of telematic presence has yet to be explored. Whenever the goal is to create a virtual common performance space between distributed musicians, the reproduction systems at both ends will influence the decision on whether incoming signals should be either filtered or dereverberated, and whether performers should be shielded by their own real signals (for example by using closed headphones) or let the virtual sound mix equally with the real sound. Performance style, latency, reproduction systems and acoustics are ultimately all interacting factors that affect the quality of performance, sometimes asymmetrically at each connected node.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] K. E. Wolf and R. Oda, "MalLo March: A Live Sonified Performance With User Interaction," in *Proceedings of the 22nd International Conference on Auditory Display - ICAD 2016*, Canberra, Australia, 2016.

[2] M. Musick, J. Turner, and T. H. Park, "INTERACTIVE AUDITORY DISPLAY OF URBAN SPATIO-ACOUSTICS," in *Proceedings of the 20th International Conference on Auditory Display (ICAD)*, New York, NY, 2014.

[3] J.-M. Jot and K.-S. Lee, "Augmented Reality Headphone Environment Rendering," in *Proceedings of the Conference on Audio for Virtual and Augmented Reality*, Los Angeles, CA, 2016.

[4] J. Lazzaro and J. Wawrzynek, "A Case for Network Musical Performance," in *Proceedings of 11th International Workshop on Network and Operation Systems Support for Digital Audio and Video*, Port Jefferson, NY, 2001.

[5] J. Braasch, B. Woodstrup, P. Oliveros, and C. Chafe, "Tele-Colonization," in *International Conference on Auditory Display*, Montreal, Quebec, 2007.

[6] A. A. Sawchuk, E. Chew, R. Zimmermann, C. Papadopoulos, and C. Kyriakakis, "From remote media immersion to distributed immersive performance," *2003 ACM SIGMM Workshop on Experiential Telepresence, ETP '03*, pp. 110–120, 2003.

[7] J. Braasch, "The Telematic Music System: Affordances for a New Instrument to Shape the Music of Tomorrow," *Contemporary Music Review*, vol. 28, no. 4-5, pp. 421–432, 2009.

[8] R. Mills, "Dislocated Sound : A Survey of Improvisation in Networked Audio Platforms," in *Proceedings of the 2010 Conference on New Interfaces for Musical Expression (NIME 2010)*, Sydney, Australia, 2010, pp. 186–191.

[9] C. Bartlette and M. Bocko, "Effect of Network Latency on Interactive Musical Performance," *Music Perception*, vol. 24, no. 1, pp. 49–62, 2006.

[10] J.-P. Caceres and A. B. Renaud, "Playing the Network : the Use of Time Delays As Musical Devices," *Procedings International Computer Music Conference*, pp. 24–29, 2008.

[11] R. Gang, S. H. Shivaswamy, S. Roessner, A. Rao, D. Headlam, and M. Bocko, "Convention Paper 8688 Audio Latency Masking in Music," in *Proceedings of the 133rd Audio Engineering Society Convention*, San Francisco, CA, 2012.

[12] C. Chafe, "Distributed internet reverberation for audio collaboration," *24th AES International Conference on Multichannel Audio, The New Reality*, pp. 1–6, 2003. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=12302

[13] N. Schuett, "The effects of latency on ensemble performance," Master's Thesis, Stanford University, 2002.

[14] E. Chew, R. Zimmermann, A. A. Sawchuk, C. Kyriakakis, C. Papadopoulos, A. R. J. François, G. Kim, A. Rizzo, and A. Volk, "Musical Interaction at a Distance: Distributed Immersive Performance," in *Proceedings of the MusicNetwork Fourth Open Workshop on Integration of Music in Multimedia Applications*, Barcelona, Spain, 2004.

[15] H. Moller, C. Jensen, D. Hammershoi, and M. Sorensen, "Design Criteria for Headphones," *Journal of the Audio Engineering Society*, vol. 43, no. 4, pp. 218–232, 1995.

[16] B. B. Boren, M. Geronazzo, F. Brinkmann, and E. Choueiri, "Coloration Metrics for Headphone Equalization," in *The 21st International Conference on Auditory Display (ICAD-2015)*, Graz, Austria, 2015.

[17] S. Marks, J. E. Estevez, and A. M. Connor, "Towards the Holodeck: Fully Immersive Virtual Reality Visualisation of Scientific and Engineering Data," *IVCNZ '14 Proceedings of the 29th International Conference on Image and Vision Computing New Zealand*, pp. 42–47, 2014.

[18] D. T. Bradley and L. M. Wang, "The effects of simple coupled volume geometry on the objective and subjective results from nonexponential decay," *The Journal of the Acoustical Society of America*, vol. 118, no. 3, p. 1480, 2005.

[19] M. Ermann and M. Johnson, "Exposure and materiality of the secondary room and its impact on the impulse response of coupled-volume concert halls," *Journal of Sound and Vibration*, vol. 284, no. 3-5, pp. 915–931, 2005.

[20] D. T. Bradley and L. M. Wang, "Quantifying the Double Slope Effect in Coupled Volume Room Systems," *Journal of Building Acoustics*, vol. 16, no. 2, pp. 105–123, 2009.

[21] C. F. Eyring, "Reverberation Time Measurements in Coupled Rooms," *Journal of the Acoustical Society of America*, vol. 3, pp. 181–206, 1931.

[22] J. S. Anderson and M. Bratos-Anderson, "Acoustic coupling effects in St Paul's Cathedral, London," *Journal of Sound and Vibration*, vol. 236, no. 2, pp. 209–225, 2000.

[23] H. Nélisse and J. Nicolas, "Characterization of a diffuse field in a reverberant room," *The Journal of the Acoustical Society of America*, vol. 101, no. 6, pp. 3517–3524, 1997. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.418313

[24] S. G. Norcross, M. Bouchard, and G. A. Soulodre, "Inverse Filtering Design Using a Minimal-Phase Target Function from Regularization," in *Proceedings of the 121st Audio Engineering Society Convention*, San Francisco, CA, 2006.

[25] F. E. Toole, *Sound Reproduction*.   Oxford, UK: Focal Press, 2008.

[26] R. Rowe and N. Rolnick, "The Technophobe and the Madman: An Internet-2 Distributed Musical," in *Proc. of the International Computer Music Conference (ICMC)*, Miami, FL, 2004.

[27] P. A. Naylor and N. D. Gaubitch, "Speech dereverberation," in *Proceedings of the 2005 International Workshop on Acoustic Echo and Noise Control (IWAENC 2005)*, Eindhoven, Netherlands, 2005.

[28] T. Nakatani, M. Miyoshi, and K. Kinoshita, "One Microphone Blind Dereverberation Based on Quasi-periodicity of Speech Signals," *Advances in Neural Information Processing Systems 16*, pp. 1417–1424, 2004.

[29] H. Kameoka, T. Nakatani, and T. Yoshioka, "ROBUST SPEECH DEREVERBERATION BASED ON NON-NEGATIVITY AND SPARSE NATURE OF SPEECH SPECTROGRAMS," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2009*, Taipei, Taiwan, 2009, pp. 45–48.

[30] M. Miyoshi and Y. Kaneda, "Inverse Filtering of Room Acoustics," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 2, pp. 145–152, 1988.

[31] M. Triki and D. T. M. Slock, "BLIND DEREVERBERATION OF A SINGLE SOURCE BASED ON MULTICHANNEL LINEAR PREDICTION," in *Proceedings of the 2005 International Workshop on Acoustic Echo and Noise Control (IWAENC 2005)*, Eindhoven, Netherlands, 2005, pp. 173–176.

[32] S. V. Amengual, M. Kob, and T. Lokki, "Investigations on stage acoustic preferences of solo trumpet players using virtual acoustics," in *Proceedings of the 14th Sound and Music Computing Conference (SMC)*, Espoo, Finland, 2017, pp. 167–174.

[33] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108*.   Audio Engineering Society, 2000.

[34] J. G. Tylka, B. B. Boren, and E. Y. Choueiri, "A Generalized Method for Fractional-Octave Smoothing of Transfer Functions that Preserves Log-Frequency Symmetry," *Journal of the Audio Engineering Society*, vol. 65, no. 3, pp. 239–245, 2017.

[35] Fischinger Timo and K. Frieler, "Influence of Virtual Room Acoustics on Choir Singing," *Psychomusicology: Music, Mind, and Brain*, vol. 25, no. 3, pp. 208–218, 2015.

[36] Z. Schärer Kalkandjiev and S. Weinzierl, "The Influence of Room Acoustics on Solo Music Performance: An Experimental Study." *Psychomusicology: Music, Mind, and Brain*, vol. 99, no. 3, pp. 433–441, 2015.

[37] B. B. Boren, M. Longair, and R. Orlowski, "Acoustic Simulation of Renaissance Venetian Churches," *Acoustics in Practice*, vol. 1, no. 2, pp. 17–28, 2013.

[38] K. Kato, K. Ueno, and K. Kawai, "Effect of Room Acoustics on Musicians' Performance. Part II: Audio Analysis of the Variations in Performed Sound Signals," *Acta Acustica United With Acustica*, vol. 101, pp. 743–759, 2015.