The World Opera: a technical, aesthetic, and philosophical introduction

Jason E. Geistweidt, PhD
Media Director, The World Opera
The VERDIONE Project
University of Tromsø
Faculty of Fine Arts
9037 Tromsø, Norway
jason@geistweidt.com

Professor Niels W. Lund
Executive Director, The World Opera
University of Tromsø
Faculty of Fine Arts
9037 Tromsø, Norway
niels.windfeld.lund@uit.no

Keywords

world opera, networked performance, mixed-reality, multi-media streaming, latency, aesthetics

Abstract

As long as there has been networked communication available, through telephone, satellite and now fiber optics, artistic experiments have made use of the network as both stage and platform for worldwide interactive live performances. Until recently the technology has not been of high enough quality to host longer, coherent works such as opera, but this is beginning to change. The World Opera is an international organization working with artist and researchers across the globe to establish a multimodal mixed-reality staging system, the World Opera House, for the production of distributed operatic works. Currently comprised of seven nodes across Northern Europe and the Americas, the World Opera and its partners are building the necessary infrastructure for real-time, interactive performances across time and space. This article reflects the authors' findings in creating and presenting distributed opera over the past five years, providing a summary of the technical, aesthetic, and philosophical discoveries/dilemmas encountered in bringing this new expressive medium to fruition. This includes a presentation of the World Opera vision, along with a summary of the Opera's developing technical infrastructure. A discussion of the aesthetic implications of mixed-reality opera follows, concluding with social and philosophical arguments for establishing the World Opera and distributed performance in general.

1. Introduction

The application of digital communication technologies to musical performance is today commonplace. The MET HDLive simulcast productions disseminate live performances from The Metropolitan Opera in New York throughout the world on a fairly regular basis. Similarly, more low-tech connections have been realized over simple consumer social networks. The YouTube orchestra reviews thousands of eager applicants each year via video auditions, discovering and, in so many ways, networking the next generation’s virtuosi who assembled this year in Sydney for a one-off performance. Both cases present a truly global undertaking in which participants are afforded

opportunities that they might not experience were it not for the technologies involved. However, neither example is accomplishing anything truly innovative. The ability to broadcast live performances globally has been with us for some time, and many orchestral auditions are still conducted by videotape. Here the processes have remained the same; only the delivery formats have changed. In contrast, The World Opera aims to change the operatic paradigm by redefining the ways in which designers, performers, and audiences *interact*.

In May 2011, Porgy stood in Stockholm while Bess, nearly 9,000 kilometers away in Palo Alto, California, declared her love for him. The accompanist, playing from the stage in California, was clear in Stockholm, and on both sides the voices blended beautifully. Porgy and Bess faced one another from their respective stages, and both audiences experienced two life-sized performers engaged in dramatic dialogue. As Bess plays coy, Porgy responds, and perhaps the true distance between the performers subconsciously amplifies the longing these two characters have for one another. This brief event encapsulates the World Opera ideal – the real-time interaction amongst individuals throughout the world, mediated by low-latency, bi-directional audio and video streaming technologies. However, it is not enough to simply provide a robust conduit for communication. We envision a network of World Opera Stages distributed throughout the globe, of a unique architecture and design, which embed the requisite technologies into the performance space itself. This is necessary in order to provide a *mixed-reality* presentation in which local performers and audiences share their space and presence with their remote counterparts virtually, thus creating a common networked space where distributed stages overlap to varying degrees.

As we clear the first decade of the twenty-first century, it is fair to say most people already experience their surroundings as a mixed-reality – juxtaposed amongst parallel worlds, consisting of one's actual embodied time and position versus the alternative locations and times mediated by the television, smartphone, or tablet. For centuries people have been drawn into virtual worlds through devices as low tech as the novel, but in recent years technology has allowed for more expressive and interactive virtual worlds. We see the inclusion of a mixed-reality dimension into the operatic toolkit as an opportunity to not only relate a story to an increasingly sophisticated networked society, but also as a way to engage those metaphors which have arisen from the use of this technology. Finally, we envision The World Opera as a natural extension of the global village movement, networking individuals and their communities together into a truly global art form. If "all the world is a stage," then the World Opera is its venue.

2. Technical Overview

The technical requirements of the World Opera are demanding, necessitating the coordination of cutting-edge audiovisual hardware and software with high-bandwidth communication networks. Additionally, these new technologies must integrate into elements of stagecraft, design, and tradition that have

evolved over the past 400 years. We are not only providing a new technological infrastructure to an already complex art form, but also providing new tools and means for expression. Thus we seek solutions that can be easily adopted by artists, technicians, designers, and audiences.

2.1 The World Opera Network

The current World Opera Network is comprised of three sites in Scandinavia – Tromsø, Stockholm, and Struer – and four sites in North America – New York, Montreal, San Francisco, and Irvine. These sites have been chosen for their proximity to major cultural resources, the availability of research networks (such as Internet2, UNINETT), and the presence of corporate and/or academic partners developing World Opera technologies.

We have chosen a high-bandwidth network solution, specifically seeking 10 Gb of capacity at each of our performance sites, if available. Higher bandwidths allow us to stream multiple channels of uncompressed audio at CD quality or better, as well as multiple streams of moderately compressed video between the various nodes of the network. Though there are interesting lower-bandwidth solutions, employing a variety of successful compression schemes, World Opera believes the demands of interactive, networked performance require a level of fidelity and low latency which higher bandwidths can more readily provide. We are also working with service providers (both public and private) to reduce the number of hops between switches within each country, providing robust routing between international gateways.

2.2 Latency

Latency is inherent within any network, and when it comes to a distributed live performance, any delay in the transmission of audio and video assets directly affects the type of interaction and production that may be attempted. Latency is the greatest hurdle to establishing a mediated performance practice. Delay is incurred at various points along the transmission chain, and results from a combination of many factors, which may be visualized simply as:

Capture Time	+	Encoding/ Compression	+	Network	+	Decoding/ Decompression	+	Display Time	=	Total Latency

Capture time describes the period required to take in and digitize a signal, that is the conversion of sound pressure waves and photons into digital audio and video streams, respectively. Additional delay is imparted as the signal is encoded for transmission, and large data streams (such as those encountered with video) necessitate compression, delaying the signal further. As assets leave the local node, they are theoretically traveling at the speed of light, the signal will slow as it traverses the myriad of switches and gateways along the network. At the remote site, the process is reversed with streams decoded, decompressed, and displayed within the space – each step incrementally increasing latency.

Keep in mind that the latency pathway offered above describes only one leg of a two-way trip, for an interactive performance requires action and reaction. For example, the video signal of a conductor's downbeat may take 150 milliseconds (ms) to reach a remote orchestra. That orchestra needs to react to the direction by producing a sound that could take 50 ms to return to the conductor in the local performance hall. Total round trip time is 200 ms, nearly a quarter of a second, and well outside the temporal range for coordinated playing (Schuett, 2002). In this example, the outgoing and incoming streams exhibit differing latency, as the process of streaming video consumes more time then that of audio.

Currently the World Opera has opted to keep audio and video streams separate, preferring to deliver assets to remote sites as fast as possible, in order to improve coordination. Thus video arrives slightly later than audio. This introduces other performances problems, but for now we believe this to be the better solution. What follows is an overview of our approach to capture, stream, and display assets for distributed productions. However, it is beyond the scope of this article to detail the intricacies of digital signal processing, compression schemes, and network optimization. For this, we direct the reader to specific studies on these topics listed at the end of this article.

2.3 Audio

The capturing of audio assets for World Opera performances is achieved primarily by placing the microphone extremely close to the source, allowing for assets to be captured without any spatial information or coloration imparted by the local space. We are working on the theory that a true-fidelity performance will present sources resonating in each respective local space. Thus, an orchestra performing in Tromsø will be captured cleanly, with little room sound, and transmitted to the remote site where it will be allowed to resonate naturally and mix more appropriately with performers in the remote acoustic. For performers on the go, wireless microphones allow freedom of movement.

Audio monitoring is provided for performers on stage if needed and speakers are placed to localize the sound appropriately (for example, the orchestra emanates from the orchestra pit). For larger ensemble works, it may be necessary to mix down signals at the source and then transmit stereo, 4-, or 8-channel sub-mixes. In addition, we can not neglect to capture sounds occurring within a space for which close microphone placement techniques may not be appropriate: audience reactions (applause) or stage sounds (the dragging of a chair across the floor). In these situations, World Opera is investigating a combination of boundary and/or hyper-cardioid microphone solutions.

World Opera utilizes CCRMA's JackTrip with Jack Audio Server (Caceras and Chafe, 2011) to both route and stream audio between performance sites. The software is easy to deploy, readily scalable, and the de facto standard for audio streaming research. We stream at a sampling rate of 44.1 KHz with a

bit depth of 16 bits/second, equivalent to CD quality sound. As previously mentioned, latency incurred for audio is quite low as compared to video, even when uncompressed. Presently, audio encoding/digitization requires about 6 ms, along with another 6 ms to decode at the remote end – 12 ms one-way and 24 ms round-trip. The physical distance the signal must travel within the network imparts further latency. As an example, network packets can travel from Tromsø to Stockholm, a 1,200 kilometer distance, in about 14 ms, or 28 ms round-trip. Thus, an audio signal originating in Tromsø will be experienced in Stockholm about 26 ms later than it actually occurred. Any response from the Stockholm side would be similarly delayed.

Additionally, there are many communications taking place during an opera, the majority of which are not be heard by performers or audience. This back channel includes conversations between stage management, lighting and sound designers, and other administrative and technical personnel. It has been our experience that this back channel is the most important audio asset during both rehearsals and production. Failure to establish a robust back channel typically results in production delays and miscommunication.

2.4 Video

The video requirements for the World Opera are the most challenging aspects of the technical setup. In addition to the latency issues previously discussed, the dynamic environment of the opera stage hinders the capture and display of video assets. Lighting, which might be appropriate for the theatre, may not be appropriate for video capture, and too much ambient lighting can easily wash out a backlit or projected display. There is a further challenge in capturing and displaying assets that have traditionally moved about the stage with relative freedom (Cooperstock, 2011a). The analogy has often been made that a World Opera production is akin to directing a live film, in which all shots must be previously storyboarded and all assets must hit their mark. It is somewhat of a chess game in which there are no second takes.

For capture of video assets, we have utilized both professional and high-end consumer video cameras. Though these devices offer excellent image quality and acceptable optics, the drawback is their use of the IEEE 1394 interface (Firewire, iLink) to bridge camera and computer, a connection that is relatively slow. With the 1394 interface, a video signal can take up to 80 ms to emerge from the camera, this is before any compression or further pre-transmission processing can take place. One option we are exploring is the use of high-end Ethernet cameras, which process images much more quickly, and due to their Ethernet connectivity, can be located some distance from the encoding computer.

As we increase the number of video streams we wish to use, it will become necessary to compress the video signals to work within available bandwidths. Compression is computationally intensive and increases latency. However total latency can be overcome if the compression scheme employed results in the faster transference of data across the network. This is possible as compression renders data into smaller packets, allowing more data to be

transferred per second. Determining the amount of compression necessary to increase transmission speeds while maintaining quality video representations is an ongoing investigation between the World Opera and our technical partners.

Video is transferred using McGill University's Ultra-Videoconferencing (UV) software (Cooperstock, 2011b). UV is a low-latency IP transport package capable of transmitting High-Definition (HD) video over the network, however, as HD consumes greater bandwidth, we currently stream with the Digital Video (DV) standard that compresses the signal within the camera. UV provides the capability of serving one stream to multiple clients and has the ability to multi-cast, though we have yet to use this function as it in not accessible in the current version. At the remote site the video is decoded, a process that typically is much faster than the original encoding. As with the cameras, consumer display technologies impart significant delay before the image is visualized. More investigation into low-latency capture and display technologies will be necessary in the future if we hope to reduce hardware-induced delay.

2.5 Other assets

A final consideration is the desire for connecting serial devices currently utilized in contemporary opera practice. The ability to sync or operate lighting and sound control boards across the network could be advantageous to coordinating distributed productions. The use of Open Sound Control (OSC) (Wright, 2003), a messaging protocol originally intended for sharing musical data, can serve this purpose well. In future, additional information, such as the location of performers on the stage, might also be captured and relayed amongst sites, creating an intelligent staging system that reacts in real time to performers' activities.

3.0 The Aesthetics of the World Opera

The World Opera is working to innovate opera production and development through the incorporation of new technologies. Though these technologies make new interactions possible, they also introduce their own peculiarities to be overcome, such as the latency previously addressed. It is precisely from this combination of innovation and limitation, however, that new aesthetics arise. As artists and technicians combine forces to explore the possibilities offered by new tools, and overcome the trade-offs encountered, they begin to develop new operatic tradition. To this point, World Opera productions have taken the form of scenes, duets and trios, out of the context of a fully staged work (Olmos, et al., 2009). As we prepare to stage our first full production, Pergolesi's *La Serva Padrona* in December 2011, and conceive of future productions, we begin to understand the implications of distributed, mixed-reality presentations for both artists and audience.

3.1 Time

In musical performance, the ability to readily synchronize events across the performance space is assumed. In a distributed presentation, this is not the case, as synchronization can vary widely, from slightly disorienting to non-performable, depending upon the latencies encountered. We must concede that certain works from the repertoire will never be presented in a distributed context successfully. However, modification of the traditional production may make this possible. For example, our presentation of *Padrona* will be a two-node production, splitting the two singing roles across the network. In the instance of recitative, a hallmark of the baroque style, tight synchronization between vocalist and continuo are required. For this reason, we are considering the presence of continuo on both stages, such that both vocalists have their own local accompaniment.

Performance also includes a visual component, and with the current infrastructure, video lags behind audio. This can be especially distressing in works that employee a conductor to synchronize large ensembles. Artist interaction is also hindered when a remote performer's mouth or facial expressions arrive out of sync with the audio. In such cases we have found that performers tend to ignore the visual component altogether, choosing to focus upon the audio component to synchronize their performance. This somewhat dulls the chemistry between performers, but we believe that this can be overcome in the rehearsal process once artists are made aware of this.

The case of a conductor coordinating a distributed performance is an interesting one. We believe it appropriate to locate a conductor where he or she may coordinate the most musicians locally. In a work like *Padrona,* this would be with the orchestra, leaving remote artists to synchronize with the networked audio. However, the top-down coordination of time, which traditionally has come from a single leader, begins to break down at this point. A conductor's reference is only to the local space. Should a the conductor ask a performer to anticipate the beat, in order to improve timing, or project more, to improve balance, the result at the remote site might be a performance that is ahead of the beat and too loud. The situation is similar with stage directors, designers, and consequently all of a production's personnel, including the audience. After all, one is only privy to the performance in front of them. Therefore, we see that distributed opera requires a communal approach to organization with multiple conductors, directors, and designers determining the best presentation for their location.

3.2 Space and motion

In a theatrical performance there is awareness among artists occupying a common performance space. Individuals have a sense of where they are in relation to their colleagues, as well as their audience. In collaboration with directors and designers, performing artists utilize space and motion to focus dramatic activity and progress narrative. In a distributed format, however, an individual's sense of place comes into question. This is true for both performer and viewer. The current challenge for World Opera is to create a

shared sense of space and presence with video technologies that are inherently static and flat.

Unlike audio, video is sensitive to line of sight issues and lighting conditions. Cameras need to be aimed in specific directions and focused upon a particular subject – projectors similarly arranged. A careless performer may obstruct a camera or projector, effectively erasing elements across the network. Further, cameras need to be in front of the subject, which makes inconspicuous placement on the local stage difficult. For purposes of display, back-projected images or large format video monitors are a better solution for discreet placement.

Our current approach is to capture individual performers in the local space and display a life-sized image at the remote site. By placing the display surfaces on the stage, among the local artists, we hope to enhance the relationship between virtual and real performer. We have had success in creating these types of mixed-reality environments through the use of the appropriately named *portal* technique. With this arrangement, a camera is aimed at a well-lit empty space, and any performer who enters this space is captured and transmitted across the network to a video display at the remote site. A benefit of this approach is that critical factors such as lighting, focus, and framing can be set ahead of time, ensuring appropriate image quality. The tradeoff is that performers are limited to predetermined areas of the stage and individuals seemingly appear and disappear as they move in and out of the camera frame. Admittedly, the technique is a bit blunt and inflexible, and we continue to search for a more dynamic solution to the placement of virtual assets in remote space.

3.3 Staging arrangements

In our discussions with designers, researchers, and artists, as well as within our limited productions, various arrangements for interconnecting local and remote spaces have been considered and attempted. We have categorized these arrangements according to the types of interaction supported and the nature of the relationships created between performers and audiences. A successful implementation will likely require a combination of these methods.

The *parallel* arrangement is perhaps the simplest to achieve, having much in common with contemporary video conferencing techniques. Here mediated performers face one another with cameras either centered on the display surface or slightly off to one side. This arrangement allows for a greater degree of interactivity, as performers are looking directly each other, as well as into one another's space. We have found that dancers especially prefer this arrangement in which spaces appear to intersect just at the projection surface. One concern, however, is the placement of audiences, which must either face the back of the local performer or be placed to the side.

Figure 1: A three-site parallel arrangement connecting (from left to right) Stockholm, Tromsø, and New York.

A *congruent* plan, which is a bit more complex, is an attempt to overlay the distributed stages upon one another, such that all elements appear to co-exist simultaneously. As opposed to the parallel setting, the congruent plan provides the illusion of remote performers entering local space, allowing artists to share a common plane, stand next to one another, and become a part of the local surroundings. This arrangement might be more preferable for audiences, but the embedding of display technologies into the performance space can cause difficulties related to the line of sight issues previously discussed.

Figure 2: Two stages from a three-sided congruent arrangement. Here the goal is to retain spatial relationships across all sites. The pianist performs from a third site, not pictured.

A third possibility for the interconnection of mediated space is the *convergent* setting, where audiences and performers share only specific elements of remote space. This is a rough compromise between the congruent and parallel arrangements and provides some insight into how spatial arrangement can directly impact narrative. Consider the possibility of an opera that is actually three, or more, operas occurring simultaneously in separate spaces. Periodically these separate works overlap, sharing a duet, a particular nuance of narrative, or a simultaneous event. In his film *Timecode* (2000), director Mike Figgis experiments with this concept of divergent/convergent narrative, allowing the viewer to follow four characters simultaneously in real-time. We wish to apply a similar experimental approach to opera, which we see resulting in a new operatic tradition and reinvigorating traditional practice.

Figure 3: A screen capture from Figgis' *Timecode* where the camera follows four intersecting narratives simultaneously in real time.

4.0 The World Opera Philosophy

The ability to present opera in this new way is a result of the technologies of the networked age. However, the *choice* to present and produce networked opera is more than a mere application of technology, rather it is based upon a philosophy and a way of seeing our society as an interconnected whole. From the Greek amphitheater to Shakespeare's Globe, from Wagner's Festspielhaus to the contemporary black box theatre, the stage has served, in one way or another, as a mirror for the surrounding society. Each époque has developed its own stage format and the World Opera stage is intended for our current époque, that of ubiquitous global interconnectivity.

We draw a comparison with the Camerata Florentine, the group of philosophers, artists, and merchants who gathered over 400 years ago to discuss the development of a new art form to mirror their contemporary world. They wished to eschew complex counterpoint and harmony in order to provide clarity to the vocal line. They insisted upon rhythms that more closely resembled natural human speech and communication.

Today we are provided with a new set of tools – Skype, Facebook, Twitter, YouTube, to mention a few – that enable real-time multi-media dialogues between distributed individuals. This is our contemporary condition, the way we communicate and how we see the world. Within the World Opera, artists, technologists, and humanists are gathering to discuss, explore, and develop new means of expression that resonate in the digital age. We wish to bring people together, share their stories, and create a truly world opera.

5.0 Acknowledgements

We would like to acknowledge the Norwegian Research Council for its support of the VERDIONE project, which has provided, in part, the means for much of the research into mixed reality performance. We would also like to thank the artistic and technical collaborators who have been indispensible to the preliminary development of the World Opera, its technology, aesthetic, and philosophy, as well as the following institutions: Stanford University, McGill University, New York University, KTH Stockholm, the University of Tromsø, and Simula Research, Oslo. Special thanks to our corporate sponsors: Bang and Olufsen, Projection Design, and DPA Microphones.

References

Caceras Juan-Pablo and Chafe, Chris (2011) 'JackTrip, A System for High-quality Audio Network Performance over the Internet', CCRMA SoundWIRE Group, http://code.google.com/p/jacktrip. Accessed 27 August 2011.

Cooperstock, Jeremy R. (2011a) 'Multimodal Telepresence Systems: Supporting Demanding Collaborative Human Activities' in *IEEE Signal Processing Magazine, special issue on Immersive Communications*, 28(1):77-86, January.

Cooperstock, Jeremy R. (2011b) 'Ultra-Videoconferencing', *McGill Ultra-Videoconferencing Research Group,* http://ultravideo.mcgill.edu. Accessed 29 August 2011.

Figgis, Mike (2000), *Timecode*, Paris: Screen Gems.

Olmos, Adriana, Brulé, M., Bouillot, N., Benovoy, M., Blum, J., Sun, H., Lund, N. W. & Cooperstock, J. R. (2009) 'Exploring the role of latency and orchestra placement on the networked performance of a distributed opera', in *12th Annual International Workshop on Presence*, Los Angeles, CA, USA, November.

Schuett, Nate (2002), 'The Effects of Latency on Ensemble Performance', Ph.D. thesis, Stanford University, California.

Wright, Matthew, Freed, A. & Momeni, A. (2003) 'Open Sound Control: State of the Art', in *Proceedings of the 2003 Conference on New Interfaces for Musical Expression (NIME-03)*, pp. 153-159, Montreal, Canada, 22-24 May.