The Telematic Learning Environment: Experiences and Outcomes Teaching with Network Audio

Diana Chester
New York University Abu Dhabi
diana.chester@nyu.edu

Abstract

The Telematic-learning environment (TLE) at New York University Abu Dhabi was created to support musical learning and collaboration, leverage network audio software to minimize latency, and connect scholars and students in the Middle East and the rest of the world. After two years of using and tweaking the TLE, the researcher has identified a variety of challenges, effective uses for the space, and learning outcomes. The following have been identified as problems. Network audio has lower latency, 25ms-200ms, than network video, 1-2sec, in the TLE. Access to high bandwidth network connections can be difficult in many places. Teachers and students must learn how to interact well with the technology in the TLE in order to be effective users of the space. Different class configurations have different impacts on teacher and student experiences. The following conclusions were found. There is a steep learning curve in the transition from VOIP solutions to the TLE. In a synchronous distance learning environment (SDLE) 50 percent more time is needed to cover the same amount of material as in a local learning environment (LLE). The relationship between teacher and student is heavily impacted by the SDLE. The telematic-learning environment (TLE) encourages teachers and students to push pedagogical limits specifically in the realm of collaboration and communication. This paper explores these issues and provides an in-depth look at the researcher's observations as both a technologist and educator in the TLE.

1. Introduction

"The transmission of audio and video in both low and high quality via the internet opens the world to otherwise impossible collaborations, a gathering of knowledge for a richer, broader musical perspective, view and exploration of the world in an expanded venue. This is the time to dream on. So we dream on! [1]" (Pauling Oliveros, *Networking Music: Low and High Tech*)

Network Audio has become a key tool in many current educational and artistic endeavors. The Telematic-Learning Environment at New York University Abu Dhabi (NYUAD) was designed in the fall of 2010 shortly after the school opened its doors. The studio was designed as a sister studio to the Telematic Studio at The Steinhardt School at New York University in New York City. The TLE was designed with the explicit purpose of supporting telematic music, defined as music that is performed live and simultaneously across geographic locations via the internet [2]. The studio was created with the intention of supporting musical collaborations, specialized music instruction for teachers and students who reside in different geographic locations, and networked musical performances (NMP), which occur when a group of musicians, located at different physical locations, interact over a network to perform as they would if located in the same room [3]. To date the studio has been used in support of credit bearing courses in music and music technology, independent music instruction between North America and the Middle East, as well as multi geographic musical and artistic collaborations.

The Telematic-learning environment is a classroom at the university that relies on network audio and video connectivity to create a specialized synchronous distance learning environment (SDLE) where low-latency allows for collaborations that would be impossible in more traditional SDLE, which often rely on VOIP or video conferencing solutions for connectivity. The TLE is set up to provide up to 8 channels of local sound amplification in conjunction with 8 channels of low-latency network audio. The studio can accommodate roughly 12 students and has a completely flexible furniture configuration with rolling stackable chairs and music stands.

2. The Technical Setup

The network video solution for the TLE is a Polycom HD videoconferencing system attached to a 56" LCD screen. This is also supplemented with Skype or IChat to accommodate participation from users who

do not have access to a video conferencing system. The network audio connection is run through JackTrip, a system for high-quality low-latency network audio performance. As JackTrip developers Cáceres and Chafe describe it, "the design achieves the highest audio quality possible, by using uncompressed linear sampling and redundancy to recover from packet loss; throughput maximization, which gets audio packets onto and off of the network as soon as the sound card can deliver them. [4]" The audio station, a large, mobile two-tiered desk, is equipped with a MacPro running OSX, a 16-channel mixing console, a digital audio interface with 8 ins and 8outs, and two reference monitors. Additionally the studio has a collection of dynamic, large diaphragm condenser, and pencil microphones, an 88-key Keyboard, microphone stands, headphones, and associated cabling.

The technical configuration of signal flow between the TLE and far-end environment is specified in diagram 1 below. The MacPro is connected to the network via a 75Mbps connection. The audio interface connects to the computer via a firewire connection, and the mixing board and audio interface are connected via analog TRS cables up to the number of input and output channels being used. On this particular audio interface there is room for 8ins and 8outs, and the default setup is for eight total channels of audio, 4ins and 4outs. The four local outgoing audio channels are routed via auxiliary outputs on the mixing board and sent through the digital audio interface into the computer via firewire and then out of the computer via JackTrip, over the 75Mbps network connection to a specified IP address on the far-end that is also running a JackTrip environment. The four remaining channels carry incoming audio via a network connection from the far-end into the local computer, patched through the digital audio interface, and are connected to four independent channel strips on the mixing board, with independent faders and EQ control.

The TLE is also configurable to support multiple channels of audio output. This allows each independent channel of audio sent from the far-end to be represented locally by an independent speaker. According to Chris Chafe, "A typical setup reproduces the ensemble's acoustical arrangement, inserting loudspeakers at the locations where a remote 'phantom' player should be. [5]" This flexibility allows for exploration of a broad range of configurations and approaches to support the learning objectives that go on within the TLE. For example, if a teacher in the TLE puts students from different environments into pairs, one from the local environment and one from the far-end, to practice or create a composition together, the audio being sent from the far-end could be connected to speakers that are placed directly next to each local student. The technologists on each side of the connection can route the incoming audio so that each student is sitting next to a speaker that plays the audio channel that directly corresponds to their partner's microphone or instrumentation. This would essentially simulate the experience of sitting next to their partner, who is in actuality on the other side of the globe.

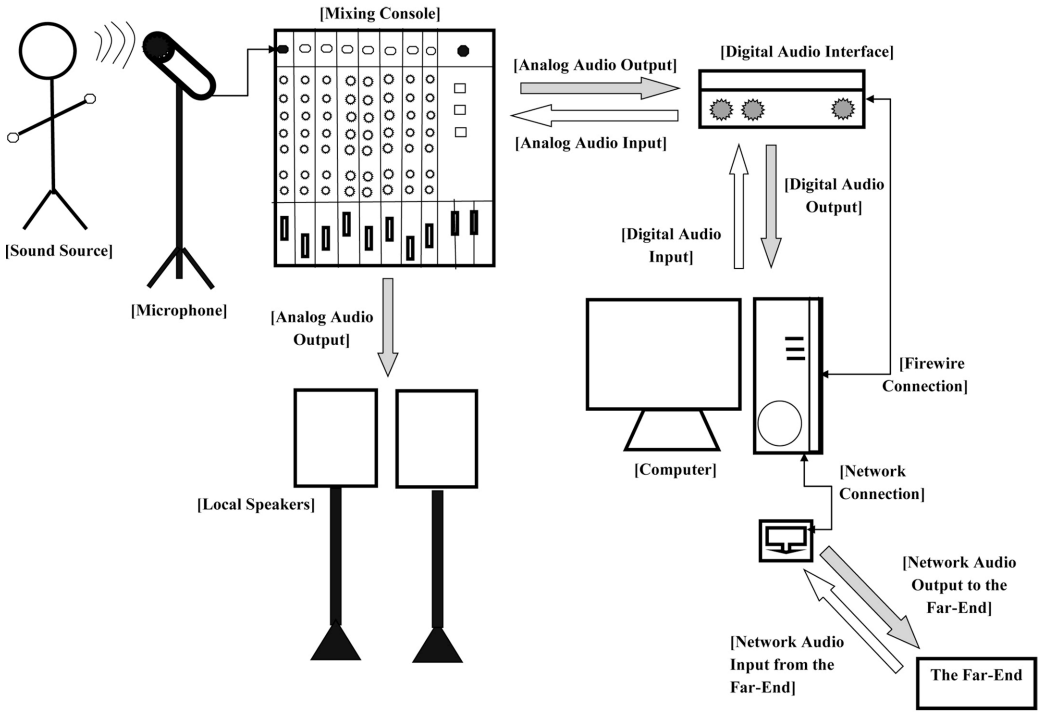

Diagram 1: Signal Flow Diagram for the Telematic-Learning Environment

3. Class Configurations

The Telematic-learning environment has thus far been used to support instruction in four distinct configurations. These configurations are as follows: (1) instructor in one geography and students in a second geography; (2) one-on-one tutoring where the instructor is in one geography and student is in a second geography; (3) instructor is in one geography and students are in both the local and the far-end geographies; and (4) instructors and students are in both the local and far-end geographies, and the class is jointly taught. The following are examples of actual course configurations that have already taken place or are currently taking place in the TLE at NYU Abu Dhabi. A Music Professor at New York University (NYU) in New York is teaching a music theory class to music majors at New York University in Abu Dhabi. A French horn instructor in New York teaches independent French horn lessons to a single student in Abu Dhabi. A guest vocal trainer in New York teaches a choral ensemble in Abu Dhabi. A global course on new music and telematic collaborations is co-taught by teachers in New York and Abu Dhabi to 6 students in Abu Dhabi and eight students in New York. In addition, the TLE has been used to bring guest speakers, performers, composers, musicians and trainers into the learning environment.

4. Technical Limitations

Maintaining consistent access to a high bandwidth network connection proves challenging in many environments. JackTrip relies on high bandwidth network connections, and the TLE relies on JackTrip for all audio connections. Hops in the network can cause static and signal interference on network audio sessions and, in some cases, results in disconnection. This is highly disruptive to the flow of a class and can often result in a huge disruption of fifteen to twenty minutes. In an SDLE this can throw the class off track entirely, forcing students and teachers to postpone the class content for a future session.

A related bandwidth limitation pertains to the number of independent audio channels being sent over the network during a given JackTrip session. JackTrip allows for variable buffer size selection to accommodate buffer glitches; however, the greater the buffer size, the greater the latency, which leads most users to maintain a lower buffer size. According to Cáceres and Chafe, buffer glitches are primarily the result of receive buffers not being sized to accommodate network jitter.[4] Due to the frequency of limitations in bandwidth and network related firewall issues, many of the JackTrip connections in the TLE are configured to run two channels of audio in and out, otherwise known as a stereo connection, in order to avoid buffer glitches that can severely disrupt or terminate the connection. The JackTrip software does have two modes, which users can activate to try to reduce the impact of buffer glitches; however, there is no way of mitigating these glitches. "Silent mode sends a packet of zeros to the process callback creating silence in the connection whenever buffer glitches occur and wavetable mode re-sends the last available packet to the process callback resulting in a synthesizer type sound whenever glitches occur in the connection."[4]

Network latency experienced in the TLE impacts the audio and video connections differently. This is due to JackTrip's low-latency system, which leads to very low latency in the audio connection – around a 25 ms-200ms lag – and much higher latency in the video connection around a 1-2 second lag. This dichotomy in the lag time of the audio and video connections creates a challenging learning environment for students and teachers where audio is heard in advance of the related video images being seen. However, the low-latency nature of the audio connection allows for musical playing and collaboration via JackTrip that would be impossible via a network audio system with a 1-2 second lag.

The Telematic-learning environment requires an intensive technical setup, making it preferable to have a dedicated space that can remain set up at all times. For some of the music lessons between Abu Dhabi and New York, the New York side does not have a dedicated TLE, and this often leads to long setup times, complications in the analog connection and routing setup (including bad hardware, network issues), and software configuration problems.

5. Technical Successes and Failures

Network Audio connections are only as strong as the network connections to which both sides have access. It is a matter of the least common denominator, whereby one strong network connection and one weak network connection equals an overall weak network connection. It is ideal to have one network audio channel per sound source, which is possible using JackTrip. Each channel of audio requires approximately 2Mbps of dedicated bandwidth. Using the example of a quartet, each instrument would have its own independent channel of audio on the local side where the live performance occurs and on the far side where the network audio of the quartet is being received. One channel of audio for each instrument of

the quartet would equal 8Mbps of dedicated bandwidth. Adding a trio on the far side would add another 3 channels of dedicated audio, at 6Mbps, which would mean that each side requires a total of 14Mbps of bandwidth capacity to support the telematic connection. Though it is ideal to have a dedicated channel for each sound source, it becomes unrealistic to do so due to the drain on bandwidth. Therefore, multiple sound sources are usually coupled together and sent over the network as one audio channel. In the NYUAD Telematic-learning environment, technicians often rely on a stereo connection of two output channels and two input channels of audio.

6. Human Limitations

Within the Telematic-learning environment, the researcher has identified three categories of human involvement and associated limitations that impact each category. The three categories are operator, user and observer of the TLE.

The operator must possess a high level of technical knowledge regarding audio engineering, the JackTrip software, network configurations, microphone placement/technique, and mixing local and network (far-end) audio. The operator must be comfortable and interested in multi-channel audio environments and up to date on new developments in network audio. The operator is usually present in the TLE when in use, trains users in the space, and should be comfortable making on the fly adjustments to the audio environment.

The user must be trained and comfortable engaging with the technology necessary to be an effective user of the TLE. This includes: (1) gaining comfortability hearing one's own voice/sounds both in the live environment and through the amplified environment; (2) ensuring that the sound source is very close to the microphone in order to send a strong signal across the connection; (3) understanding how to avoid microphone feedback created by proximity to speakers; and (4) becoming familiar with proper SDLE etiquette to maximize effective communication. It can be disconcerting to talk to someone who does not appear to be looking at you; to mitigate this common problem it is critical that the screen and camera are placed close together in both the local and far-end environments.

The observer needs to be comfortable seeing key collaborators only through a screen and hearing amplified versions of sound rather than source sounds. Additionally, observers are often not able to contribute, as observing in the TLE is very different from observing in an LLE. In a local learning environment, the observer is equipped to contribute but may not participate in the same way as the user, though they have all necessary tools to do so. In the TLE the observer does not have a microphone, which means they cannot participate in the conversation with the far-end. The observer is usually not captured by the camera as the camera has a tight crop on the users, making the observer invisible to the far-end. Based on configuration of the TLE and spatial constraints, the observer will likely have a poor view of the far-end making it difficult to follow the far-end communication and interactions.

7. Human Successes and Failures

In a synchronized distance learning environment, where there are students in two geographies, it is challenging for students to truly connect with the students on the far-end in the same way they connect with their peers in the LLE. This makes sense given the students' ability to communicate with their local peers in a broader range of dimensions and modalities.

One challenge of the TLE is student disengagement with class content, as it does not take into account as broad a range of learning styles as an LLE. This is due to two main factors: (1) the technology used in the TLE can be less inclusive of different learning styles; and (2) not all traditional teaching methods translate well from the LLE to the TLE, e.g. the use of whiteboards for sharing notes with students. Additionally the technology itself can be a disengaging factor for students and teachers, as there is more of a barrier to vocally participating in the TLE than there is in an LLE.

The user's experience of their sonic environment impacts how integrated the local and far-end environments feel. Ideally the technologist is able to mix the sound well so that local and far-end audio sound similar in their overall decibel levels. This is a challenging feat given the variety of variables in play including number of users, types of instrumentation, output levels, and sonic fluctuations. Especially in a musical collaboration context, major changes in tone, velocity, and volume can create sudden sonic implications for the experience of students and teachers on the far-end. These sudden changes in sound quality create a scenario where students and teachers frequently become responsible for self-mixing their sonic environment until the technologists can reconfigure the sound on both sides of the connection. Though reconfiguring the sound may only take seconds to accomplish, the frequency and suddenness of major changes in the soundscape can be jarring and create major distractions in learning.

8. Conclusions: The TLE vs. Traditional SDLEs

The learning curve in using the telematic studio is very steep relative to using a VOIP tool like Skype, or Google voice. Technologists who operate the telematic studio must be familiar with analog audio, complex audio engineering, digital audio interfaces, microphone technique, video conferencing and internet video tools and have a familiarity with basic networking concepts. There are often only a few people in an organization who posses the correct combination of skills to be successful engineers in the TLE. This means that when the technologists are not available, the TLE cannot be utilized. The researcher has found that this often leads to guerilla approaches that sacrifice low-latency control and the ability to work in a multi-channel environment.

The capacity for multiple channels of audio to be sent and received via JackTrip differentiates the TLE from other SDLE's whose audio connections are limited to a stereo input/output. The multiple audio channels ensure clarity and quality from each sound source, as each is independently controllable. In the example of independent music instruction, students and instructors are in the TLE in their respective geographies. They have one microphone for conversation and often a secondary microphone that captures audio from their instrument or in the case of a direct input instrument (e.g. keyboard or electric guitar) the second channel of audio is the direct input. The technologist on each side of the connection has the ability to control the quality, clarity, equalization and relative decibel level of the local audio and the network audio being sent from the far-end. The technologists send their two channels of audio over the network connection, uncompressed, to the far-end. When the far-end channels of audio are received, the technologist will mix the two far-end channels (voice and instrument) so that they sound sonically appropriate with the voice and instrumentation in the local TLE. This granular control means that users are able to collaborate on music and sonic creations in an environment where audio quality, clarity and control are paramount, and contribute to keeping an academic focus on creativity and exploration, not on the technology or the logistical elements of the connection environment.

8.1. Experiencing Time in the SDLEs

A two-hour class that involves two geographies and a network audio and video connection is equivalent to roughly a one and a half hour class held in a local environment. When teaching within SDLE's, a teacher requires roughly 50 percent more class time to cover the same material they could otherwise cover in an LLE. This is due to a number of factors both technical and practical that impact SDLEs and not LLEs. These factors include technical setup and troubleshooting, network latency, technology in the learning environment as a distraction, and the need to define new modes of communication that are appropriate for the SDLE.

Though time is factored in for setup and troubleshooting ahead of the class start time, it is often the case that technology fails or takes longer to work than one can anticipate. A small glitch in an otherwise satisfactory setup can easily take fifteen to twenty minutes to resolve, delaying the class start and often frustrating students and teachers who are not able to connect with each other during this down time. When technology does not work in an LLE, (e.g. the projector does not successfully connect to the laptop inhibiting a teacher from sharing a presentation), the class goes on until the technical issue is resolved. In the SDLE, there is no class if the technology does not work correctly.

Network latency impacts the flow of communication and creates a break in conversation that people are not used to in real time environments. This often results in people talking over one another, staggered video images and long pauses in conversation. Latency can be challenging even for seasoned users of network audio and video tools. For classes and instruction held in the telematic-learning environment, latency presents a new set of challenges because the amount of latency present in the network audio connection, roughly 25-200 milliseconds, is severely less than that present in the network video connection, roughly 1.5 seconds. This results in students and teachers hearing one another speak or play before they can see those same actions on the screen in front of them. The difference in latency is especially confusing when trying to conduct a musical collaboration or when relying on visual cues.

The presence of technology in the TLE is unavoidable and greets students and teachers at every turn. A brief orientation on how to properly engage with the TLE includes training on how close one must be when talking or playing into a microphone, where the camera is located in the room so that one can look directly into it, and to instruct participants that they will be heard before seen. The technology is critical and exciting and enables collaborations that would otherwise be impossible over distance; however, it can also be a barrier to effective communication or an unwanted presence in a learning space.

In an LLE teachers and students rely on many non-verbal indicators and gestures to communicate, while the SDLE requires them to redefine and create new modes of communication. Something as simple as being able to see if students have a particular book in front of them allows a teacher to ask students to open up to a specific page in that book. In the SDLE the teacher may be able to see that there are books on the desk, but likely cannot determine which books they are, requiring an added layer of communication, where the teacher asks the students if they have that book. The students must respond in order for the class to move forward. This added layer of communication is time consuming and requires teachers and students to recalibrate the way they are used to functioning within the classroom. A second example of this is a teacher playing a song from their laptop for their students on the far-end. In an LLE the teacher can discern the volume of the song and adjust it appropriately for the class. In an SDLE the teacher must rely on the technologist in the far-end classroom to act as a translator and to adjust the volume appropriately, when often the technologist has no reference for the particular piece of music and what it should sound like. To further complicate matters it can take the teacher 20-30 seconds into playing the musical piece to know if the music is successfully playing on the far-end. This is because it will often take that long for the students to recognize that there is a problem, since they may not be certain if/when the piece was started. This is an example of how the back and forth communication, often unnecessary in an LLE, can be a complicating and time consuming factor in an SDLE.

8.2. How SDLEs Impact Teacher/Student Interaction

Synchronous Distance Learning Environments impact the discourse and relationship between teachers and students. For a class configuration where students are in one geography and the teacher is in a second geography, communication outside of class time can be complicated by time zone differences, the inability to hold office hours, and cost implications. Since scheduling time outside of class time can be challenging, teachers are often faced with the decision of spending large amounts of time coordinating and holding telephone or Skype sessions with each student or contacting students only by e-mail. Often interactions outside of the class time are limited to e-mail only, which is an inadequate means of communication for delving into curricular content.

It is very important to avoid class configurations where a teacher is in one geography teaching a class of students in that same local geography as well as students in a second geography. When this configuration occurs, it can create a learning environment in which a teacher is more attuned to the communication styles and collaborative spirit of the students in their local geography than those on the far-end. Students on the far-end may often feel disconnected from a learning community that is being created in the geography where the teacher is located. Since it is inherently easier for most people to connect in person than via network audio and video, this class configuration privileges the students who are located in the teacher's geography. The far-end students' communication and comfort level with the teacher can be negatively impacted by their perception of the heightened communication and ease of rapport between the teacher and the students in the teacher's local geography.

In a class configuration where there is a teacher and a group of students on both the local and far-end, student/teacher interaction will be stronger between the teacher and students who are located in the same geography. Students and teachers in the same geography will be able to interact less formally, with fewer technical barriers, and have side bar conversations before, after and even during the class. This same informal communication would be very challenging between students who are in one geography and a teacher in a second geography. It is common in this configuration for the local teacher to become the conduit through which students' conversations and ideas are shared with the far-end. At times, the teacher becomes a moderator and helps to insert the voices of the local students into a conversation on the far-end when students find it difficult to do this on their own. This is due primarily to the fact that students have more familiarity, comfort, and experience in the LLE than they do in the SDLE.

Due to this same phenomenon, class configurations where the teacher is in one geography and students are in a second geography seem to lack the same back-and-forth discourse found in the LLE. In this configuration it is common that those students who are outspoken and have more dominant personalities are most comfortable interacting with the teacher on the far-end. Sometimes one such student takes on the role played by the local teacher in the previous example and represents or repeats the ideas of their quieter peers to the teacher on the far-end. It is important for teachers to recognize when this happens and to try and engage the quieter students in the class. The technology can serve as a barrier that deters some students from participating in the class in a way that is audible to the far-end, but may be inaudible to their peers in the LLE.

8.3. The Telematic-Learning Environment and How it Supports Global Learning

The ability to leverage teaching resources from the Middle East, North America, and anywhere else in the world exponentially expands learning opportunities, teaching opportunities and the ability to explore new arenas of pedagogy linked to explorations made possible via network audio. According to the *AES White Paper: Best Practices in Network Audio*, the use of network audio for musical collaboration allows for the exploration of pedagogy, styles and techniques not previously available and creates the potential for greater pedagogical accomplishment in a specified time period [6]. The TLE exposes students to musicians, scholars, peers and collaborators from around the globe.

One challenge inherent in relying on network audio to create a learning environment is that there are many places in the world that have limited access to good network connectivity. Since the TLE and JackTrip software specifically require high bandwidth capacity, limited access to a network connection with a high bandwidth capacity could result in the need for a collaborator to find a local university or government facility that is equipped with the appropriate resources in order to participate. In some cases the technical requirements can become exclusive of people participating in the TLE.

The technology in the TLE is currently bringing educators from around the world to the Middle East and from the Middle East to other parts of the world. This is able to happen virtually without the cost of plane tickets or ever needing to leave one's geography. At present collaborations taking place via NYUAD's TLE are focused primarily between North America, the United Arab Emirates, and East Asia. However since the technology can support collaborations between any and all geographies, provided technical resources and network capacity are available, the TLE will undoubtedly grow and expand to connect students and educators across the globe.

9. References

[1] Oliveros, Pauline*, Networked Music: Low and High Tech,* Contemporary Music Review*,* Vol 28. Nos. 4/5, August/October 2009, pp. 433-435.

[2] Oliveros, P, Weaver, S, Dresser, M, Pitcher, J, Braasch, J, Chafe, C, *Telematic Music: Six Perspectives*, Online supplement to LMJ 19, 2009, MIT Press Journals.

[3] Lazzaro, J.;Wawrzynek, J, *A Case for Netowrk Musical Performance*, NOSSDAV '01: Proceedings of the 11th international workshop on Network and operating systems support for digital audio and video. ACM Press New York, NY, USA, 2001, pp. 157–166.

[4] Cáceres, Juan-Pablo and Chafe, Chris, *Jacktrip: Under the Hood of an Engine for Network Audio*, Journal of New Music Research, Vol. 39, No. 3, 2010, pp. 183-187.

[5] Chafe, Chris, *Tapping into the Internet as an Acoustical/Musical Medium*, Contemporary Music Review, Vol. 28, Nos. 4/5 August/October 2009, pp. 413-420.

[6] Boulliot, Nicolas, *AES White Paper: Best Practices in Network Audio,* J. Audio Eng. Soc., Vol. 57, No. 9, September 2009.