Converge:

An Omni-Biographical Composition

Jieun Oh

CCRMA, Department of Music, Stanford University, USA

jieun5@ccrma.stanford.edu

http://ccrma.stanford.edu/~jieun5/

Ge Wang

CCRMA, Department of Music, Stanford University, USA

ge@ccrma.stanford.edu

http://ccrma.stanford.edu/~ge/

Converge is a project that leverages ubiquity of mobile computing to capture compositional materials from our daily lives, which become re-contextualized during a performance through a sonification and visualization of the past. Participants in the composition record sounds, take pictures, and submit text descriptions using mobile phones to capture their life moments. During the performance of *Converge*, these timestamped and location-tagged snapshots are brought together, converging to the present time and place. By accompanying us at all times wherever we go, mobile computing serves a critical role in this composition, generating a musical experience that is personal and powerful. This paper details the overarching motivation, technical implementation, and performance context of *Converge*, viewed as a unique example of a composition realized by a new model of computing that is pervasive, both in the physical and affective sense.

With a surge in smartphone penetration level and the accompanying increase in the exploration of interaction techniques on mobile devices, we are motivated to experiment with new social models of music making that can be made feasible through mobile technology.

Gaye, Holmquist, Behrendt, and Tanaka described in 2006 how mobile music could enable novel forms of musical experience by taking advantage of changes in social and geographical context [3]. Indeed, mobile technology has recently been actively explored in computer music research as a platform for audio synthesis and control. Tanaka and Gemeinboeck pioneered in the exploration of location properties of mobile music and media artworks [9][10]. Greg Schiemer's *Pocket Gamelan* [8] is one of the first examples of using mobile phone as a physical musical instrument. Levin used audience members' mobile phones to choreograph spatialized ringtones [4]. For a more complete description of related works in mobile music, see Wang, Essl, and Penttinen (2008) [12].

More recently, structured ensembles using mobile phones as primary musical instruments were founded to explore new possibilities in music making [5][6]. These ensembles have experimented with interaction techniques that rely on gestures (using multi-touch, accelerometer, and gyroscope), location (using compass and GPS), and networking (using OSC) in conjunction with real time audio synthesis and graphics rendering, demonstrating how transforming mobile phones into physical instruments has become an achievable reality [1]. Essl and Rohs provide a theoretical analysis of sensor capabilities in mobile phones, and show how they can facilitate interactive performances on mobile devices [2].

A desirable next-step in the exploration of mobile music is focusing more on social and geographical elements of performance, beyond the physical interactions. Such musical experiences may manifest partly on-device, and partly in back-end "cloud computing" servers, and seek to connect users through music making. The iPhone Ocarina [11] is an early experiment that creates a social experience by allowing users to listen to one another; other mobile interfaces and experiences have been designed to put together a concert for audience participation using social mobile computing [7].

increasing freedom	Spatial Dimension	Temporal Dimension
	audience participation model performance participation by audience members who are present in the concert space	***fully synchronous model*** all actions occur during a window of scheduled performance time and are immediately audible
	centralized model performance participation by people not physically present in the concert space	***quasi-synchronous model*** actions preceding a set performance "start-time" get incorporated during the performance
	decentralized peer-to-peer model completely distributed performance without a centralized performance location	***asynchronous model*** no set performance time, and actions are merged at a later time acc. to predefined rules

Figure 1. Conception of Possible Social Models for Mobile Music Making

We hope to continue experimenting with the different social models for mobile music making, based on the degree of freedom in the spatial- and temporal-dimensions exercised by the performance, as illustrated in Figure 1.

Vision

The *Converge* project involves participation from people from geographically diverse locations, as described under "audience participation model" and "centralized model," and following a "quasi-asynchronous model" in the temporal dimension (according to Figure 1). It tries to investigate the question, "what kind of new music can we make, now that we have mobile, physical interaction devices that are connected in geographically diverse locations?"

Converge constructs an audio-visual composition from a set of autobiographical moments that are collected by a crowd, using cloud-mediated mobile applications. The source materials are collected over a large interval of time prior to the performance, such that the collected events originate from the lives of the participants, occurring asynchronously in various geographical locations.

This compositional process is an example of 'ubiquitous music' at work, unveiling a new potential for music making made possible by an invisible fabric of mobile technology. Mark Weiser, in his landmark paper "The Computer for the 21st Century," described how "the most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it" [13]. *Converge* may serve as a musical attestation to Weiser's remark.

In *Converge*, we decided to capture autobiographical moments using the medium of text description, photographed image, and recorded sound. The performance itself pays special attention to the temporal and geographical information of the originating sources by presenting them relative to the present moment (that is, the time and location of the performance setting). By re-contextualizing the past experienced by the participants through a variety of expressive medium, the resulting composition can be regarded as an "omni"-biography of all of its participants.

In implementing the composition, we keep in mind the following overarching objectives:

- Social Paradigm: Explore spatially distributed and temporally asynchronous models of music making

- Methodology: Leverage crowds and clouds, through use of mobile computing

- Composition: Attain coherence in composition amidst diversity of source materials.

- Performance: Balance improvisation (low-level presentation of performed sources) and scoring (high-level planning of performance trajectory)
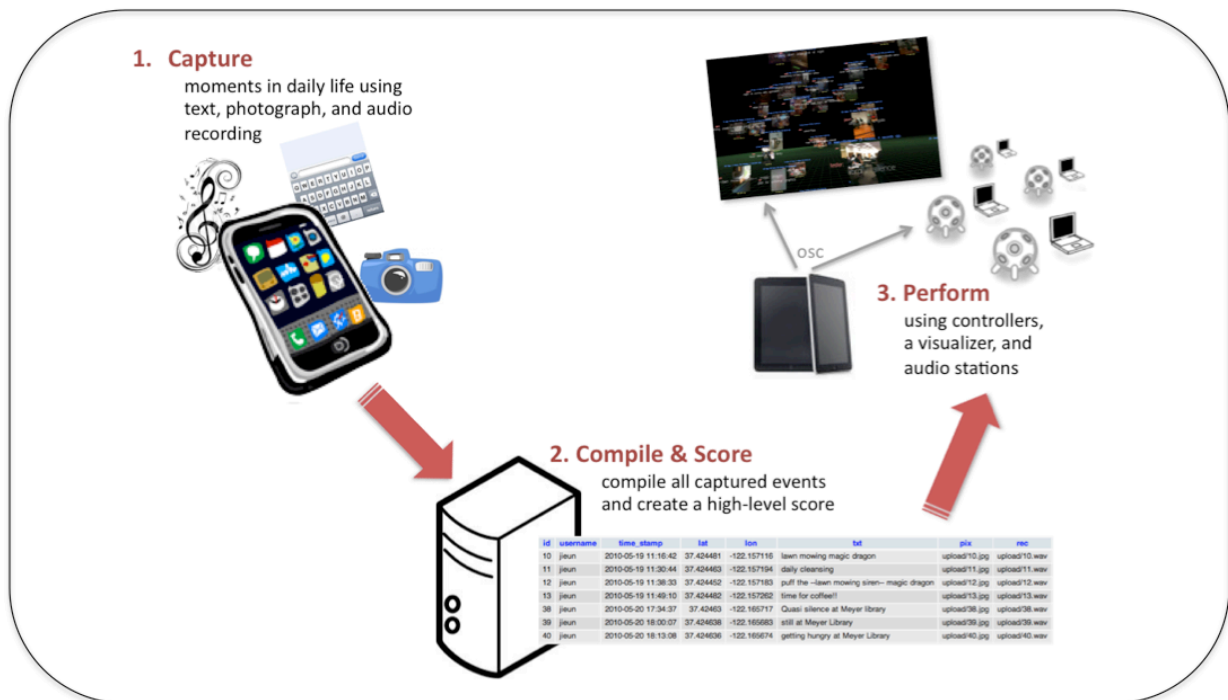

Figure 2. Three major stages in *Converge*

2

Implementation

This section summarizes the major stages involved in crafting the composition, from source material collection to performance. Figure 2 summarizes the entire process.

Acquiring Compositional Elements

The first stage involves acquiring compositional elements, and we experimented with various methods to collect them. A common goal, irrespective of the method, was to leverage mobile computing to allow participants to readily capture moments in their daily lives that they find meaningful.

A custom iPhone application, named *Convergence*, was one method used to allow users to capture data and upload them to a server.
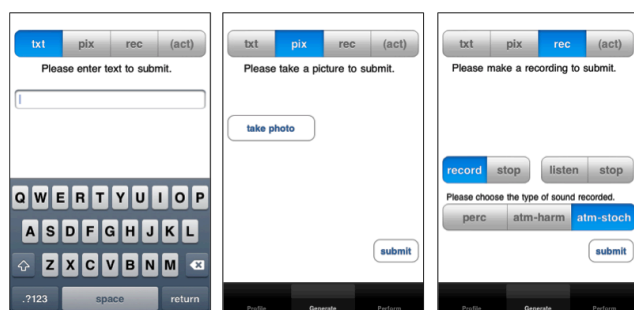

Figure 3. Text, Picture, Audio Recording Submission

As shown in Figure 3, segmented control user interface at the top of the view allows users to select the submission type: text, picture, or recording. With the text option, users can type in a short description of the moment they are experiencing in the provided textbox. For the picture option, users can take a photograph of what they are currently seeing, using the mobile phone's built-in camera. Finally, the recording option provides a simple interface to start and stop recording sound, and to listen to the recording just made. Captured sound can be of anything that the users are currently hearing; upon recording, users may characterize the sound as percussive, atmosphere harmonic, or atmosphere stochastic.

Camera and ***Voice Memos***, which are two default applications on iOS devices for taking photographs and making audio recordings, respectively, are another ways of acquiring the source materials. Both programs have an email feature to share the captured data. Users are instructed to send an email with the subject "convergeThis" and include in the email body a text description of the photograph and audio recording.

Unlike the first method of using a single custom application, this method requires using multiple applications to submit all three types of data. Users must additionally provide their username and a description of their current location, as *Voice Memos* did not have a built-in functionality for obtaining the latitude and longitude of current location. Despite such inconvenience, using these default iOS applications was the method of choice for *Converge 2.0* (described in "Performances" section below), which involved submissions from a large group of participants, mainly because the *Convergence* app was not available for public download from the Apple App Store.

Twitter (http://twitter.com) was the final method we tried for collecting data. For this method, participants are asked to use third party *Twitter* apps that enable posting photos and audio recordings. This method is theoretically appealing because posting one's status on *Twitter* using mobile device is a casual social activity, something people already voluntarily perform on a daily (perhaps even hourly) basis.

A tweet submission needs to include a hash tag, #convergeThis, to make it possible for us to later search for the submissions in the compilation stage. Also, participants are asked to enable location feature of *Twitter*, which automatically pulls user's current location.

As of 2010, several sites and services such as *twitpic* (http://twitpic.com) and *yfrog* (http://yfrog.com) offered easy uploading of pictures to *Twitter*. Unfortunately, however, fewer viable options existed for audio. Moreover, not many third party applications included the location functionality at the time, such that the submission entry would be missing an important piece of information needed for the performance. Finally, *Twitter*'s search feature for hash tags only returned recent posts such that older submissions would not be retrieved from the search query. For these various reasons, this method turned out to be unappealing for our purpose.

Compiling and Scoring Collected Elements

The second stage comprises of compiling and scoring the collected elements. Upon completing the acquisition of source materials, we create a CSV file consisting of entries of all usable submissions. Then, the CSV file and all image files are saved in the *visualizer*, and all audio files are saved in the *audio stations*. The entry IDs, which are also used as names for the submitted images and recordings, are loaded onto the *iPad controls* for use in the performance. (The italicized components used in a performance are described in the "Performance Technology" subsection below.)

From the pool of all usable submissions, we pick out about 20-30 entries to plan out the narrative trajectory of a performance, targeted for around 10-15 minutes in total duration. The score has an arch-like trajectory: events (submitted entries) are triggered in a roughly

id	username	time_stamp	lat	lon	txt	pix	rec	soundtype
10	jieun	2010-05-19 11:16:42	37.424481	-122.157116	lawn mowing magic dragon	upload/10.jpg	upload/10.wav	0
11	jieun	2010-05-19 11:30:44	37.424463	-122.157194	daily cleansing	upload/11.jpg	upload/11.wav	0
12	jieun	2010-05-19 11:38:33	37.424452	-122.157183	puff the --lawn mowing siren-- magic dragon	upload/12.jpg	upload/12.wav	0
13	jieun	2010-05-19 11:49:10	37.424482	-122.157262	time for coffee!!	upload/13.jpg	upload/13.wav	0
38	jieun	2010-05-20 17:34:37	37.42463	-122.165717	Quasi silence at Meyer library	upload/38.jpg	upload/38.wav	0
39	jieun	2010-05-20 18:00:07	37.424638	-122.165683	still at Meyer Library	upload/39.jpg	upload/39.wav	0
40	jieun	2010-05-20 18:13:08	37.424636	-122.165674	getting hungry at Meyer Library	upload/40.jpg	upload/40.wav	3
41	tester	2010-05-20 18:18:32	37.42463	-122.165711	library silence	upload/41.jpg	upload/41.wav	0
44	jieun	2010-05-23 14:25:08	37.42289	-122.169235	Ge typing	upload/44.jpg	upload/44.wav	1
45	ge	2010-05-23 15:30:37	37.421326	-122.17263	working on this piece	upload/45.jpg	upload/45.wav	2
46	ge	2010-05-23 16:49:26	37.423387	-122.170621	coffeeeeeeeeeeee	upload/46.jpg	upload/46.wav	0
47	oh	2010-05-23 16:55:37	37.42289	-122.169226	data collecting on my iPhone	upload/47.jpg	upload/47.wav	0
48	ge	2010-05-23 23:06:31	37.416373	-122.147158	cruising down page mill at night	upload/48.jpg	upload/48.wav	0
49	jieun	2010-05-26 09:07:31	37.424598	-122.1541	morning bird songs	upload/49.jpg	upload/49.wav	2
50	jieun	2010-05-26 09:13:44	37.424275	-122.156378	birds and airplanes- things that fly!	upload/50.jpg	upload/50.wav	2
51	ge	2010-05-26 17:13:07	37.423387	-122.170621	plan for converge graphics	upload/51.jpg	upload/51.wav	3

Figure 4. Sample Database Table

chronological order until the climax, at which point a galactic vortex of images converge into a dense singularity, and all triggered instances of audio intensify to a loud, rich texture. Following a brief moment of complete darkness and silence, the composition's denouement is played out as an organized rebirth of visuals and sound, illustrated as a bright explosion of light and a delicate sounding of a single-frequency sine wave. Figure 5 shows a sample score used in one performance, with event IDs (in blue) and their accompanying audio (left column) and visual (right column) summaries.
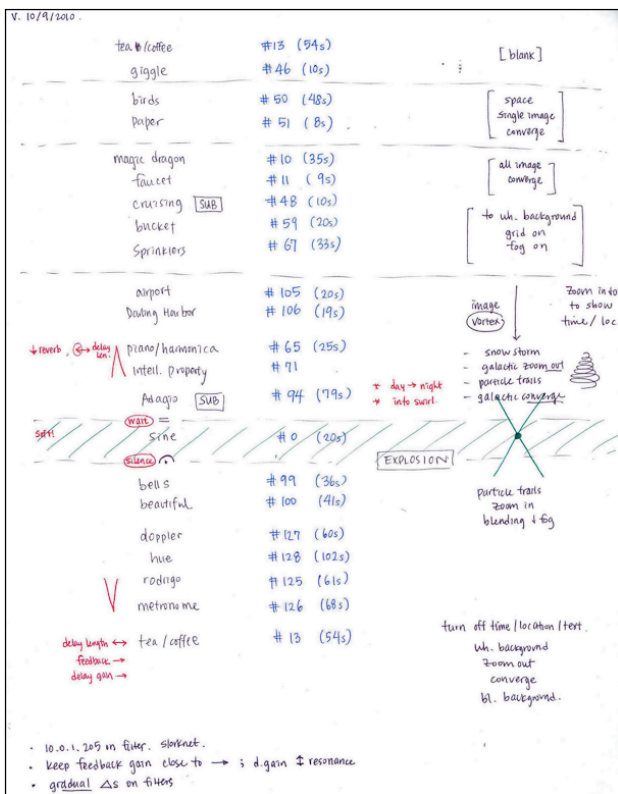

Figure 5. Sample Score

Performance Technology

The third and final stage is a live performance of the composition. The major components of the technology used to perform *Converge* are controllers, a visualizer, and audio stations. The controllers are software interfaces used to trigger audio-visual events. The visualizer is responsible for dynamically generating 3D graphics, which are projected on a large screen. Finally, the audio stations are audio synthesis units running ChucK and outputting sound. Two performers—one person on the controllers and the other on the visualizer—are needed for the performance.

The controllers are interfaces for sending OSC messages to the visualizer and the audio stations, thereby tightly connecting the three components of the performance technology. The controllers are implemented as iPad applications because they offer flexible interface programming for touch-based controls.

The controllers are used to trigger source events (using *cv-source*), as well as to apply filtering effects (using *cv-filter)*. Figure 6 shows prototype interfaces.

In *cv-source*, the performer selects the data source and specifies channels for one or more audio stations. Upon touching the "Trigger" button, OSC messages are sent to (1) selected audio stations, and optionally also to (2) *cv-filter* and (3) the visualizer. Consequently, the chosen audio stations begin playing on the specified channels the sound file corresponding to the source selection. If the "OSC to Filter" switch is on, then the "Trigger" button also sends an OSC message to the *cv-filter* controller with the precise stations- and channels-settings used in the trigger, which get dynamically loaded to *cv-filter*'s UIPickerView; selecting the latest item in this UIPicker-View, in effect, applies filtering effects to the same audio stations and channels that were just triggered using *cv-source*. Finally, an OSC message gets sent to the visual-

4

Figure 6. Controllers Interface: *cv-source* (left) and *cv-filter* (right)

izer, to trigger the image and text visualizations that correspond to the chosen source event.

In *cv-filter*, the performer can set various filter parameters, for real-time audio effects. Parameters for control include overall gain, overall reverb, comb filter delay length (set either on a continuously spectrum in terms of the number of milliseconds of delay, or on a discrete selector in terms of the resulting resonant frequency), feedback amount, and delay gain. Similar to in *cv-source*, the performer selects audio stations and channels to which the parameter settings should be applied. An OSC message gets sent to selected stations for audio digital filtering, and optionally also to the visualizer for image distortion effects.

The visualizer is a desktop application implemented using OpenGL and C++. The program dynamically generates visual contents based on the input CSV file, image files, and keyboard controls. The resulting graphics is projected on a screen during a performance.

In the visualizer, each source event is encapsulated as a CvNode, which stores the following properties associated with the event: the username of the participant who submitted the event, the timestamp on the submission, latitude-longitude for the submission, the text description of the event, and the photograph submission. Each of these properties of CvNode can be made visible or hidden based on artistic goals during the performance.

This information is presented in a three-dimensional space in a way that contextualizes the event's time and location relative to the present moment: we calculate the distance between the performance hall and the latitude-longitude of the event, and displaying the text, "k kilometers away," over the image. Similarly, we calculate the difference in time between the present moment and the timestamp on the submissions, displaying the text "d days, h hours, m minutes, s seconds ago." The text is updated every second so that audience sees the event moving away into history as they are experiencing the per-

Figure 7. Converge Visualizer Output

formance. This has served to generate a powerful experience for the audience, eliciting feelings of time travel and nostalgia.

The performer on visualizer has a great control over the presentation of the three-dimensional space and the various elements residing in it. For instance, the background color can change from black (Figure 8 left) to white (Figure 8, middle), and the materials can be presented with a grid (Figure 8, left), or without (Figure 8, middle). Other global effects controlled by the keyboard include fog rendering and animated trails on moving images. Elements can scatter, converge, and crumple in space; and they can form a vortex whose spiral gravity and speed can also be controlled. The camera can be zoomed in and out. All of these controls are slewed to produce smooth, organic transitions.

Certain events in the visualizer are handled by the controllers: *cv-source* triggers a specific CvNode corresponding to the source data, bringing it into the center of view, and *cv-filter* controller distorts the image with the change of filter parameters.

Audio stations are either laptop orchestra stations (full or chamber-sized) with hemispherical speaker arrays; mobile devices (iPhones or iPads) that are attached to speaker gloves; or multi-channel speakers. We use ChucK with OSC listeners to process audio. Upon receiving an OSC message from the *cv-source* controller, audio stations trigger the specified audio file on specific channels. Low-pass filtering and reverb effects can be optionally applied to the triggered audio, based on the OSC messages sent from the *cv-filter* controller.

Performances

Converge has been performed a number of times using different source-acquisition methods and performance techniques in multiple performance settings. This section highlights unique characteristics from each of the major performances of *Converge*, and explains how the project has evolved. Each iteration resulted in diversification of the source materials, eventually reaching a maximum event-to-event geographical distance of over 10,000 km and temporal range of over six months.

Converge 1.0

In a set of performances categorized under *Converge 1.0*, the authors collected all of the source materials using the *Convergence* application. We tried out various stage setups to optimize the performance experience.

SLOrk Spring Concert. The premiere performance was part of a larger laptop orchestra concert held in June 2010. Events making up the composition were local
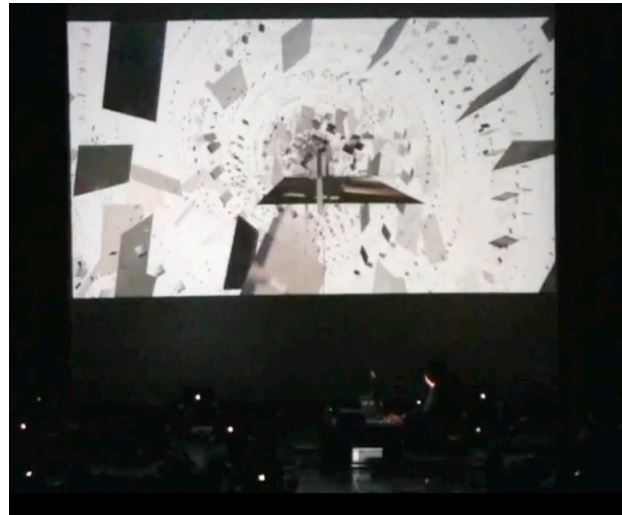

Figure 8. Performance in SLOrk Spring Concert

(within 20-mile radius of the performance location) and spanned a relatively short time range of about three weeks. The performance took advantage of a large stage space and performance space by using all eighteen laptop orchestra stations. Consequently, the resulting soundscape was quite dense and rich. The visualization, too, was projected on a very large screen in a completely dark space, generating an immersive experience, much like in a theatre.

Figure 8 shows a snapshot of the performance. A video documentation of this performance can be found at http://www.youtube.com/watch?v=dhA2IusZubk.

Mobile Concert at NIME. The second performance of *Converge* occurred about two weeks following the premiere, as part of the Mobile Concert of *New Interfaces for Musical Expression Conference 2010*, held in Sydney, Australia.

The source materials used in the performance included not only the ones collected in preparation for the premiere performance, but also newly captured events at the conference site. As a result, we were able to achieve great geographical diversity of source events: the performance involved triggering events that occurred thousands of kilometers away, as well as within 10km of the performance location.

Figure 9 shows a snapshot of the performance. As conveyed in the picture, the projection on glass windows in a partially-lit space generated an atmosphere that was unfortunately not as immersive as was the first performance. Another major difference between this and the prior performance was in the audio stations: instead of using laptop orchestra stations, we used six iPads running ChucK to process and output audio. The decision for this was made out of a practical need, as travelling

6

Figure 9. Performance in NIME Mobile Concert

abroad with multiple laptop orchestra stations and hemi-spherical speakers was challenging.

During the performance, six iPads were scattered out in the performance space. We amplified audio outputs by connecting speaker gloves to the iPads. Overall, the resulting soundscape was more intimate and spatialized.

Converge 2.0

Converge 2.0 is an attempt at a "crowdsourced" composition, and was performed as part of "MoPhO 2.0: Audience Participation Concert" held at CCRMA Stage, Stanford University, on November 18, 2010. Instructions for participation were described through a webpage, https://ccrma.stanford.edu/~jieun5/converge. A total of twenty-three submissions were made over a period of ten days, by students, faculty, and non-university affiliated individuals, mostly from the local area but also from across the nation. Combined with the existing pool of material collected by the authors in the previous months, the new submission allowed us to have a more diverse pool of events from which to craft our composition.

Evaluation

This project is just a beginning in our exploration of how mobile technology can enable new social models for music making, an experimentation in spatial-distribution and temporal-asynchrony in compositional elements. Converge encourages participation in the compositional process by anyone from virtually anywhere and anytime. Other technology, such as mobile controllers that send OSC messages over the network, stations that synthesize audio, and graphics rendering that dynamically generates scenes in tight synchrony with the audio, all made it possible to craft an evocative audio-visual experience.

Audience Response

Audience members commented on how they felt nostalgic during the performance. The animation of ticking timestamps was a constant reminder that these moments are continuously moving away in time. Even though most (if not all) of the events that made up the composition were not of their own, people mentioned how it felt as though they had actually lived those moments themselves before; people vicariously experienced each other's past moments, making the composition "*omni*"-*biographical*.

Following a performance at Foo Camp, one person described how "In a weekend of surprises it was the absolute highlight for me and many others who were in a similar state of awe." Another commented, "the idea of capturing an experiential moment and eliciting them as musical memory is a really powerful concept - beautiful in its simplicity and with huge potential for developing and reinterpreting at any type of scope."

Composers' Reflections

Upon each iteration of *Converge*, especially during the "compiling and scoring" stage of the compositional process, the authors were emotionally affected by the massively autobiographical nature of the piece. Creating a

Date	Location	Setting	Description
June 2, 2010	Dinkelspiel Auditorium, Stanford University	Stanford Laptop Orchestra Spring Concert 2010	Premiere. Using full laptop orchestra (~18 stations) as audio stations
June 18, 2010	University of Technology, Sydney, Australia	New Interfaces for Musical Expression 2010 Conference	Using mobile devices as audio stations. Geographical diversity in source materials
June 26, 2010	O'Reilly Media HQ, Sebastopol, California	Foo Camp 2010, organized by O'Reilly Media	Using chamber-size laptop orchestra (6 stations). Greater geographical diversity and temporal range
November 18, 2010	CCRMA Stage, Stanford University	Stanford Mobile Phone Orchestra Audience Participation Concert	Using chamber-size laptop orchestra. Greater participation/ collaboration

Table 1. Summary of Major *Converge* Performances

narrative to be used for the upcoming performance required recalling and reusing materials used in the preceding performances. The piece, in effect, became a documentation of its own evolution, with a high proportion of events captured from the stressful and exciting moments of performance preparation.

Future Directions

We hope to continue developing *Converge* with our initial vision in mind, enriching its social and collaborative elements. While the existing version does achieve crowd-based social participation in the initial source collection stage of the composition, we hope to also enable such collaboration in the compiling and scoring stage as well as during the performance stage.

For instance, participants could offer suggestions, through a web interface, on the ordering of events in scoring the piece, or tag descriptors on the submitted events to allow automation in scoring elements to a desired narrative trajectory.

Collaborative participation during the performance would allow audience members to have partial control over when to trigger which event and apply audio-visual effects. Taking this idea a step further, dynamically processing new incoming data from participants while the performance is in progress would enable a rich performance experience. Such a design would allow even people who are not physically present in the performance space to impact the outcome of the performance.

References

[1] Bryan, N., Herrera, J., Oh, J., and Wang, G. (2010) "MoMu: A Mobile Music Toolkit." In *Proceedings of the International Conference on New interfaces for Musical Expression.* Sydney.

[2] Essl, G. and Rohs, M. (2009) "Interactivity for Mobile Music-Making." *Organised Sound*. 14, no. 2: 197-207.

[3] Gaye, L., Holmquist, E., Behrendt, F., and Tanaka, A. (2006) "Mobile Music Technology: Report on an Emerging Community." In *Proceedings of the International Conference on New Instruments for Musical Expression.* Paris.

[4] Levin, G. *Dialtones – a telesymphony.* http://www.flong.com/projects/telesymphony/ Accessed January 20, 2011

[5] The Michigan Mobile Phone Ensemble. http://mopho.eecs.umich.edu/ Accessed January 20, 2011

[6] Oh, J., Herrera, J., Bryan, N., Dahl, L., and Wang, G. (2010) "Evolving the Mobile Phone Orchestra." In *Proceedings of the International Conferences on New Interfaces for Musical Expression*. Sydney.

[7] Oh, J., and Wang, G. (2011) "Audience-Participation Techniques Based on Social Mobile Computing." In *Proceedings of the International Computer Music Conference.* Huddersfield.

[8] G. Schiemer and M. Havryliv. (2006) "Pocket Gamelan: Tuneable trajectories for flying sources in Mandala 3 and Mandala 4." In *Proceedings of the International Conference on New Interfaces for Musical Expression.* Paris.

[9] Tanaka, A. and Gemeinboeck, P. (2006) "A framework for spatial interaction in locative media." In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Paris.

[10] Tanaka, A. and Gemeinboeck, P. (2006) *net d'erive.* Project web page. http://www.ataut.net/site/Net-Derive Accessed January 20, 2011.

[11] Wang, G. (2009) "Designing Smule's iPhone Ocarina." In *Proceedings of the International Conference on New Interfaces for Musical Expression. Pittsburgh.*

[12] Wang, G, Essl, G., and Pentinnen, H. (2008) "Do Mobile Phones Dream of Electric Orchestras?" In *Proceedings of the International Computer Music Conference*. Belfast.

[13] Weiser, Mark. (1991) "The Computer for the 21st Century." *Scientific American.* 265, no. 3: 94-104.