# A Framework for Spatial Interaction in Locative Media

Atau Tanaka
Sony Computer Science Laboratory Paris
6 rue Amyot
F-75005 Paris FRANCE
+33-1-4408-0512
tanaka@csl.sony.fr

Petra Gemeinboeck
Key Centre of Design Computing and Cognition
University of Sydney
NSW 2006 AUSTRALIA
+61-2-9351-3030
petra@arch.usyd.edu.au

## ABSTRACT
This paper presents the concepts and techniques used in a family of location based multimedia works. The paper has three main sections: 1.) to describe the architecture of an audio-visual hardware/software framework we have developed for the realization of a series of locative media artworks, 2.) to discuss the theoretical and conceptual underpinnings motivating the design of the technical framework, and 3.) to elicit from this, fundamental issues and questions that can be generalized and applicable to the growing practice of locative media.

## Keywords
Mobile music, urban fiction, locative media.

## 1. INTRODUCTION
This paper presents the concepts and techniques created to realize a body of location sensitive mobile media artworks. They are installation works extending beyond the confines of a gallery, to include the urban environment. Conceived for participative use from advanced mobile telephones, they call upon techniques from interactive music applied to new contexts that include the interplay of sound and image, an exchange between participants in the gallery and participants in the streets, and the creation of an abstract narrative from audiovisual media captured on multiple mobile devices.

The paper is organized as follows. In section 2 we describe related works, both by the present authors, as well as others in the field. Section 3 introduces the notions of spatial interaction and locative media that form the conceptual basis of the work and section 4 outlines a typical usage scenario. Section 5 describes a generalized system architecture. The system is ultimately a framework that can be adapted to a series of works in the area. This framework is able to service various embodiments, including the scenario described in Section 3. Section 6 closes with discussion of the first installation work created using the framework.

## 2. RELATED WORK
The authors have each separately produced prior work that leads up the present collaboration. Tanaka has published on and demonstrated a collaborative mobile music system [18]. Gemeinboeck has exhibited and published installation works in which abstract threads of memory interweave mediated and physical spaces [5].

Related work in the area of mobile music systems can be seen in the projects presented at international events like the Mobile Music Workshop series [11]. Recently, new fields of creative works known as *locative media* are gaining interest. Work in this area is represented in research consortia such as Open PLAN [8], and in festivals such as PixelAche [13]. One of the themes for the 2006 edition of the International Symposium of Electronic Art (ISEA 2006) is the *Interactive City* [10].

Recent artistic inteventions in the field of *locative media* include the annotation of the physical with virtual marks or narratives [14] or the revelation of patterns by tracing the participants' movements through the environment [16]. They tend to take a documentalist approach to integrating geography as a component in artistic practice.

## 3. CONCEPTS
### 3.1 Spatial Interaction
The work seeks to establish a connection across and coherence between two distinct types of spaces – urban space, and a gallery exhibition space. The latter is typically characterized as a "white box", controlled environment with artificial lighting to present cultural artefacts. Everything is done to allow concentration and minimize distraction in a gallery. An urban environment, on the other hand is an uncontrollable multiplicity of independent activity. Typified by the term, "concrete jungle", it is characterized by chaotic activity, and multifarious events, indications, and communication that simultaneously make demands on our attention.

### 3.2 Locative Media
The deployment of mobile, networked, location-aware computing devices, involving participants in mapping processes, social networking or artistic interventions is often associated with the emergent field of locative media. Considering geographical space to be a canvas that allows the inscription of personal narratives, desires, and memories, offers a powerful instrument for communities to (co-)author their environment and to collectively organize and share such subjective experiences. There is, however, a double-edge to the attempt of inversing power strategies of remote control, as the technologies affording these collective location-based interventions paradoxically operate upon the same plane as surveillance. Its reliance on positional precision and location-based context critically link locative media to the arena of cartography and its dominant practices of mapping. And, in Irit Rogoff's words, "mapping as a cultural, political and epistemological activity is deeply imbricated in nations' narratives of their own formation" [15]. This ambivalent notion, of course, bears the potential of opening up a *third space* [1], a space for intervention, in which these power relations can be investigated and negotiated.

## 3.3 Geography as Musical Interface

This paper seeks to describe location-based techniques in the context of interactive music and multimedia. Interactive music has until now primarily focused on the use of sensor systems as the underlying technologies used to build novel interfaces and new musical instruments. Here we describe the use of geographic localization in this light. As the system is multi-user, it is not only location that enters the system as musical parameters, but the community interactions that take place amongst the users of the system. While prior work by the authors has demonstrated the use of social dynamic to generate musical re-mixes, here we describe a generalized framework that can support the creation of a body of work that is location aware, multi-user, and multimedia.

# 4. SCENARIO

## 4.1 The Players

There are two types of participants corresponding to the two spaces the work inhabits: the outdoor mobile environment, and the indoor gallery environment. Players outdoors are armed with specialized mobile phones loaned out to participants. The indoor participants not only experience the resulting audio/visual projections in the gallery, but interact with the piece and the other participants through touchscreen interfaces on pedestals in the gallery space.

By spanning two disparate environments with a single system, we create a 'third space' that is spanned by three mobile participants, three gallery participants and server. The interconnections correlate, process and communicate actions in each domain of this shared space, resulting in the community media dynamic that is the output of the system.

## 4.2 Mobile Participants

A series of advanced mobile telephones are made available to gallery visitors who are encouraged to undertake a 'mission' in the neighborhood surrounding the gallery. The mobile phone functions as a music listening device with Walkman headphones, and with its built-in video camera, as a portable video camera.

Depending on the specific manifestation, or use of the framework in a specific piece, the *motivation* may take on different forms. This may be game-like, with specific targets as in a treasure hunt, more social, as in orchestrating personal encounters through sound/image avatar mediated representation, or be more abstract or even aimless. The framework supports all these modes, the specific embodiment will define the participants' *tasks* through a kind of community choreography.

## 4.3 Gallery Participants

The gallery space is arranged to be a sort of "mission command central". The recomposed collage of incoming images from each mobile camera and processed and mixed sounds are projected on a large-scale screen and in the sound system. They display the ongoing weaving process in relation to the path that is invisible for the participants outside the gallery.

The gallery participants are not just passive viewers, but are able to interact with the mobile participants via three small touchscreen interfaces. Made available on pedestals in front of the main projection screen, each touchscreen corresponds to one mobile phone out in the field. The touchscreen display is similar to the mobile phone display, but larger, having greater functionality, and giving a higher level view of the system. While the display follows the camera image received from its corresponding mobile phone (surveying the surveyor…), it also shows the relation of the current path and the one it seeks to follow (the trail of a previous participant). Allowing the gallery participant some level of control, they can choose to highlight or obscure relations before they are interpreted and sent to the participant outside. These manipulations not only contribute to the progress of the *mission*, but are also displayed on the large screen and sound system.

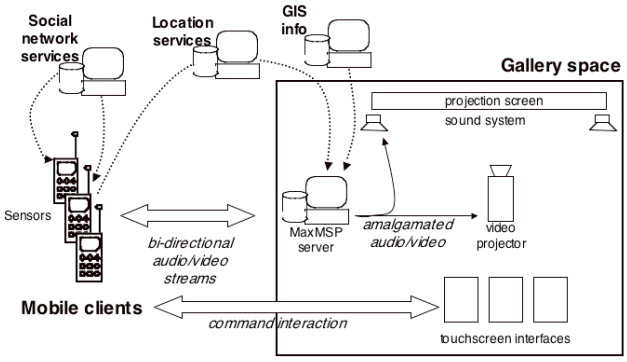


**Figure 1. System Overview**

## 4.4 Data Aggregation

The system server aggregates the different audio video sources from the mobile devices, and processes them according to control information coming from the different participants and various external sources.

Mobile participants are able to interact in a crude fashion using the 5-way control buttons on their mobile phone. In a more advanced implementation, sensors onboard the mobile devices can give more fluid information on the gesture and movement of the mobile user. While these various control data can directly affect the audio/visual display local to their device, their primary use is to be recovered by the system as a series of XML messages that alter that participant's part within the aggregated whole on the server.

Likewise, gallery participants actions on the touchscreens serve not only to send instructions to their corresponding mobile user, but also are sent up to the server to contribute to global media processing. The exact processes and mappings depend on the specific implementation.

External control sources include geographic location of each mobile device. While existing systems like the Global Positioning System (GPS) allow accurate location tracking, they necessitate additional hardware. In the current framework, we make use of mobile phone' GSM cell antenna signal reception strength mapped to mobile operators' databases of antenna location to deduce an approximate location. A Location Based Services (LBS) server provides the gateway between the measured antenna signal, the operator database, and the framework server.

Other external information includes Geographic Information Systems (GIS). This allows recovering demographic information correlated to mobile client position. Socio-cultural embodiments of the framework make use of this data channel to extend the *third space* afforded by the framework to include socio-economic data that is location specific.

## 4.5 System Output

Based on aggregated media and associated control information, the system server amalgamates the disparate media elements, processing, combining, mixing, distorting them according to rules and mappings defined in each specific embodiment.

The result is a series of audio/video streams sent out to several destinations. The destinations include returns to the individual mobile devices and in-gallery touch screen displays. The primary destination is the main video and sound projection in the gallery, representing the sum output of the system. The media being streamed out to the mobile devices and touchscreens can be individually tailored, or may mirror the main output.

Sound output is rendered based on the interaction elements coming in from the control channels. Two types of sound are employed: pre-composed musical tracks, and live environmental recordings. Each mobile user may have selected a music track as their "soundtrack" for their walk. Meanwhile, the microphone on the mobile phone serves to provide live sound capture of the outdoor environment. These elements are correlated to similar elements from the other users, as well as history of previous users experiences to create a collage mix. Depending on the state of the user's sensors or location, as well as those of the others, the final sound output will be an evolving mix of deconstructed musical source elements injected with environmental noises, actively panning and substituting elements from other users and other moments while preserving a coherent sonic flow.

# 5. ARCHITECTURE

The technical infrastructure of the piece supports both classical client-server architectures as well as peer-to-peer like end-to-end communication. The project server is situated in the gallery space. There are three kinds of server side functions: audio/video (a/v) input/output (i/o), message services, and location services. There are two types of clients: the mobile devices and the small touchscreens in the gallery.

## 5.1 Clients

### *5.1.1 Mobile Terminal*

The client is a mobile telephone with multimedia capabilities. It has a network connection over 3G (or UMTS) packet switching mobile data infrastructures with a bandwidth throughput of up to 256 kbps.

The terminal is able to operate as an audio and video capture device, streaming live audio-video streams up the wireless network to the server in the gallery. It is also a sound/image display device, with MP3 audio decoding played over standard Walkman headphones, and with a 240x320 color graphics display. The client is able to read audio and video data streams from the network to display in real time. A custom Java program was developed, and has been downloaded to the phone's memory prior to the exhibition.

The user interacts with the system via a 5-way joystick-button system and two programmable soft-buttons. The phone folds switchblade style, concealing the numeric telephone keypad, but leaving the camera, display, audio jacks and interface buttons apparent. This configuration allows the phone to lose it's telephone associations and be used as a nomadic a/v device. (Fig. 2.)

While the primary mode of communication for the mobile client is with the project server, there is the possibility of direct peer communication amongst the mobile devices. A profiling system is used to construct a social networking map when each mobile device joins the system. The social network defines the situations and conditions in which mobile devices would enter into direct communication. This in conjunction with the location services could for example aid two mobile users to cross paths if nearby.

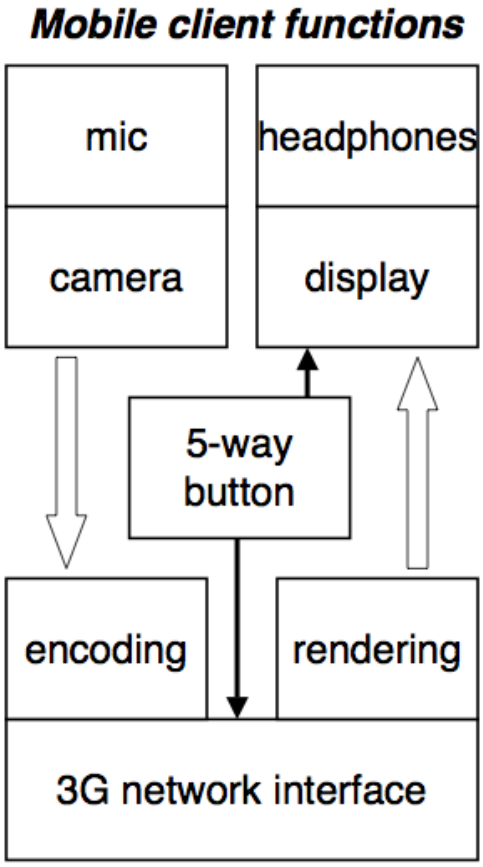


**Figure 2. Mobile Client Architecture**

### *5.1.2 In-gallery Touch Screen*

The stationary clients are 12.5cm diagonal 768x1024 pixel handheld computers. They are connected via 802.11g (WiFi) wireless networking to the server. One touchscreen interface corresponds to each mobile phone out in the field. The user is able to manipulate the graphic user interface (GUI) on the touchscreen with his fingers. These in-gallery client devices serve as mediators between the mobile terminals and the final gallery audio/video projections.

## 5.2 Servers

### *5.2.1 A/V Server*

The server is programmed in MaxMSP/Jitter with several ancillary components. It is able to receive parallel audio/video upstreams from each of the three mobile phones.

The server has three distinct a/v output destinations: the mobile clients, the gallery clients, and the main projection display.

Audio to the mobile clients is sent up to an Icecast server [9] running on the same machine that relays the audio stream to the mobile phones as an MP3 live stream.

The incoming video from the mobile phones is mirrored to each corresponding in-gallery client over the LAN via http. The video stream is consolidated and encoded as an H.263 streamed over HTTP to the mobile clients.

The monitor output as well as audio output of the server machine are connected to a local projection system in the gallery. This is the final combined output of the system, sent to a video projector and sound system.

### *5.2.2 Messaging Server*

It also receives control input resulting from the mobile user's actions on the telephone buttons, as well as the gallery participants' activities on the touch screens. These control message enter as OSC messages translated to XML over UDP and TCP socket connections.

### *5.2.3 Location Server*

Location based services are provided as a middleware layer server side between the mobile phone operator and the gallery. Each mobile client can be tracked geographically to a precision determined by the distance between two cell antennas. This location data is aggregated at the gallery server as one of the visual/musical parameters.

### *5.2.4 Social Network Server*

A user profile system allows for recording of user preferences and user-specific metadata. While on commercial systems, this kind of data would typically be used for personalization to feed recommendation engines, in the case of the present framework, the data is used as a means through which different users may enter into direct contact. This allows bootstrapping of a social network overlay to the system where principles from social computing are used to make connections between direct acquaintances, friends of friends, and users of similar profiles. This provides the peer-to-peer communications channel that runs parallel to the client-server media architecture predominant in the framework.

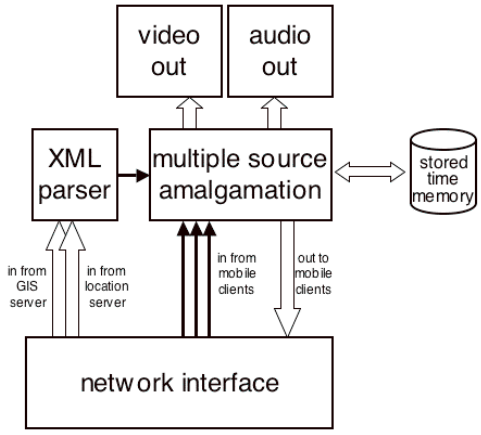


**Figure 3. Server Architecture**

# 6. DISCUSSION

## 6.1 Idiomatic Locative Media

We have described the architecture and general usage scenario of a hardware/software framework that is not specific to one artwork, but that should be useful in the creation of a series of musical or audio-visual works that seek to tap the dynamic of locative media. The question arises about the true generality of the framework, and idiosyncrasies that may influence the resulting works.

The motivation was to create a framework detached from a specific creative embodiment much in the way that a musical instrument has an independent existence and identity from the pieces that are composed for it. A musical instrument, while separate from a work for that instrument, however, does exhibit specificities and idiosyncrasies. Composers who successfully exploit these idiosyncrasies, we say, have found the *voice* of the instrument, and have written music *idiomatic* for that instrument [20]. An instrument can foster a repertoire, or a body of work, that while stylistically diverse, remain identified with that instrument.

We sought a similar goal with the present framework. We did not want the framework to be tied down to needs or peculiarities of a single piece. At the same time, the construction of the framework represents the views of the present authors on issues of locative media. By building a *locative media instrument* of sorts, we hope that others may benefit from the efforts in creating the system to realize other, related works. The notion that we have created a platform even for ourselves to realize more than one piece work is attractive. The resulting pieces, whether done by us or by other composers and artists, may have the indelible identity that associates it with the framework described here, but from the idiomatic composition perspective, this is not seen as a negative quality.

In the following sections we describe some of the theoretical underpinnings that motivated us to construct the framework in the way we did.

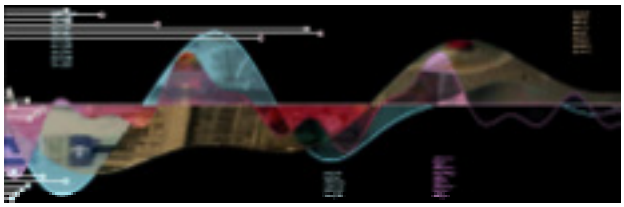

**Figure 4. Example of generated imagery**


**Figure 5. Image displayed on mobile client**

## 6.2 Memory and the Situationist Dérive

The participants' mission is to uncover the traces of earlier participants, that is to say that they are using their mobile instruments to look for and to follow the invisible trails of now past participants and the traces captured along these paths. The scenario reminds of a notion of Benjamin's *flaneur*, in which they put themselves in the shoes of someone else in order to emplace themselves in the city. [2] Only that the observed other is no longer present. The city in the 'eyes' of the device then composes and decomposes itself in relation to the invisible path. As both current and stored positions and images are compared on the server, each participant, similar to the game 'Hot and Cold.' receives indications about their closeness to the path. These visual and acoustic indications, however, can be deceiving as the result of these relational processes can be obscured or manipulated by the gallery participants. Participants, however, don't only leave and receive traces but can also

choose to conspire and to co-navigate themselves through the neighborhood. The path reveals itself through puzzle pieces that appear on the screen, weaving threads of the past into the urban present.

## 6.3 Surveillance

The digital revolution has helped make possible the surveillance society as predicted by Orwell and Huxley [12, 7]. While governments put in place networked capability to survey with advanced image analysis techniques all in the name of security, few today decry Orwell's 1984 and Big Brother except by anecdote. Why?

Perhaps one explanation is that as governments have gained surveillance powers, so has the public. The omnipresence of consumer devices camcorders has created a revolution in *grassroots journalism*. Beginning with the Rodney King video in 1991 denouncing police brutality [6], amateur video has often been the first camera on the spot, be it for the tsunami in Southeast Asia, or the crash of the Concorde [3].

The question of balancing of power, or tension between powers, respectively, and its potential inversion of observer and observed create the second dynamic we wished to put in place in the work.

## 6.4 Interrelations

The server, storing and processing the trails and traces of the participants, of course, appears to have the ultimate 'knowledge' and thus control over the relations of the individual participants. The tension unfolds when the gallery visitors use their 'control power' to manipulate the sets of relations or the mobile participants, in turn, conspire with one another.

The process of detecting and negotiating the invisible paths not only relies on the instrument's positions but also deploys pattern and image recognition strategies. Analogous to the participants' experience of seeking the invisible trail, the underlying control process likewise becomes a matter of relations and associations, rather than a one-to-one correspondence of geographical positions (coordinates).

# 7. CONCLUSIONS

We have described a framework for the creation of musical and audio-visual artworks that extend outside of the concert hall or gallery space to embrace and include the surrounding city. The system is conceived as a semi-generalized framework, much in the way of a musical instrument. We have described the qualities and structure of the instrument, and have defended the architectural decisions by a theoretical stance with respect to location-based interaction.

While the conceptual structure calls upon cultural theory, the modes of interaction draw directly upon techniques that are current in the fields of interactive music, media art, and social computing. This provides the composer or artist the possibility to extend their craft beyond the confines of traditional cultural venues to include the dynamics of the urban environment.

# 8. ACKNOWLEDGMENTS

We would like to acknowledge the participation of Clicmobile S.A. for their gracious contribution of location and social network services. This work has been funded in part by a grant from the RIAM, the French Minister of Industry.

# 9. REFERENCES

[1] Bhabha, Homi, K. *The Location of Culture*. Routledge, London, 1994.

[2] Bull, M. "The World According to Sound: Investigating the world of Walkman users". In *New Media and Society*, 3(2), 2001, 170-197.

[3] "Concorde: What Went Wrong?" In *BBC News* September 5, 2000.

[4] Gaye, L., Mazé, R., and Holmquist, L. E. "Sonic City: The Urban Environment as a Musical Interface." In *Proceedings of International Conference on New Interfaces for Musical Expression (NIME03)*, Montreal, 2003.

[5] Gemeinboeck, P. "Impossible Geographies of Belonging". In *ACM Multimedia 2005 Proceedings*, November 6–11, 2005, Singapore. ACM Press, New York, NY, 2005.

[6] Gooding-Williams, R. *Reading Rodney King / Reading Urban Uprising*. Routledge, Inc., London, 1993.

[7] Huxley, A. *Brave New World*. Harper & Bros., New York, 1932.

[8] Open Plan. http://www.open-plan.org/

[9] Icecast. http://www.icecast.org

[10] International Symposium on Electronic Art, San Jose, 2006. http://www.isea2006.sjsu.edu

[11] International Workshop on Mobile Music Technologies. http://www.viktoria.se/fal/events/mobilemusic/

[12] Orwell, G. *1984*. Harcourt Brace Jovanovich, New York, 1949.

[13] PixelAche Festival. http://www.pixelache.ac

[14] West, N. "Urban Tapestries: the Spatial and the Social on your Mobile." In Bart Lootsma. (ed.) *La ville à nu / The naked city. Archilab 2004.* 2004.

[15] Rogoff, I. *terra infirma: geography's visual culture*. Routledge, London, 2000.

[16] Rueb, T. *Choreography of Everyday Movement*, 2002. http://mitpress2.mit.edu/e-journals/LEA/GALLERY/locative/choreography/index.htm

[17] Sannella, M. J. *Constraint Satisfaction and Debugging for Interactive User Interfaces.* Ph.D. Thesis, University of Washington, Seattle, WA, 1994.

[18] Tanaka, A., Tokui, N., Momeni, A. "Facilitating Collective Musical Creativity." In *ACM Multimedia 2005 Proceedings*, November 6–11, 2005, Singapore. ACM Press, New York, NY, 2005.

[19] Tanaka, A. "Mobile Music Making." In *Proceedings of International Conference on New Interfaces for Musical Expression (NIME04)*, Hamamatsu, 2004.

[20] Tanaka, A. "Interaction, Agency, Experience, and the Future of Music." In Brown, B., O'Hara, K. (Eds.) *Consuming Music Together: Social and Collaborative Aspects of Music Consumption Technologies.* Computer Supported Cooperative Work (CSCW) Vol. 35. Springer, Dordrecht. 2006.