Between Place and Interface: Designing Situated Sound for the iPhone

Anders Fagerjord
Department of Media and Communication, University of Oslo

Introduction

During 2010, led by the example of Apple’s iPhone, advanced telephones with touch screen, Internet capabilities, and GPS receivers became widespread. Social network sites such as Facebook, Twitter, Gowalla and Foursquare allowed users to share their physical location. The authoring of text, imagery, or sounds for specific locations—what we might call situated texts— is quickly being added to the list of digital competences. This article reports on our experience in creating an application (or “app”) for the iPhone explaining music history in Rome, using situated sound in the form of music and spoken commentary.

The idea for the app was simple: a tourist in Rome walks into one of Rome’s many churches. Her iPhone detects which church she is in, and plays music written for that room, along with spoken commentary explaining the music and the church’s role in music history. Our focus was on creating a well-composed experience inside the church, much like a radio program or a magazine article, rather than a new kind of interface or an application for social interaction online.

Following a research by design approach, we tried to answer these questions: What are the affordances of sound in locative media? How can one design an interesting and enjoyable location-based learning experience in audio? Is multimodal composition for a specific place different from less specific kinds of authoring?

Our application has, like many Web sites, both navigation and “content”, that is, short radio-like essays on music history, but unlike most Web sites, the

user does not navigate a metaphorical cyberspace, but the streets of Rome’s central neighborhoods.

Content is not a very specific label. We have tried to create an example of a possible future genre (as suggested by Liestøl 2006, p. 263). We found that when we took care to select music and write commentaries adapted to the special mood in each church, we were able to position the listener in what Lev Manovich (2003, p. 2) called an augmented space, a space not completely present in the church, focusing on a time that is neither current nor historical.

I begin by describing our app briefly so the reader may understand what I am discussing, before accounting for the design method. Then I discuss how the design has taken into consideration the small size of the iPhone, the user’s situation in a busy city, and the affordances of sound in a specific location.

Design description

A basic interface element in the iPhone is the simple list (Apple 2008. pp.). Users move from list to list, animated as a movement from left to right. Going “back” to a previous list is conceptualized as a leftward movement. All churches are available from two main views: a plain list of churches, and a map (see Figure 1).

![Figure 1: List and map view](image-url)
The map shows a view of the historical centre of Rome, and the list is sorted alphabetically, but both views may be reordered from the user's location. When the “location” button is pressed, the map repositions, and shows her position with a marker. In list view, the list of churches is sorted by proximity, from closest to most distant.

Selecting a church in the list or on the map will bring forward its “Outside view” (Figure 2). The outside view guides the user to the church’s entrance, and has a short description of the church and the music offered at that place. All churches are not large landmarks, and some have entrances that are hard to find. We found that an exterior image, taken from the direction where tourists are likely to approach the place, could be of help.

**Inside view**

From the outside view, the user may press the “enter church” button and walk into the church in question. The inside view (Figure 3) is black, to signify a different mode, inside a dim church. It carries a photo of the church's interior and short introduction to the church, mostly to give an idea of what the interior looks like, in case the user is not there. The main feature is a set of playback controls. When the user is ready to listen, she can press play, and listen to music composed for this particular room, and a spoken commentary.
pointing out how the structure and aesthetics of the music matches that of the architecture and art.

The playback controls are similar to those of Apple's iPod app, to ensure familiarity and ease of use. As in the iPod app, the inside view may be “flipped”: A touch of the button in the upper right will change to a track list via a "flip" transition. The track list shows other pieces of music with commentary created for the church. When satisfied, the user can use the “back” buttons to go back to the list or map view.

This current version is the result of an iterative research by design process, which we will account for now.

Method: research by design and synthetic-analytic method

We have followed the approach Birger Sevaldson described as Research by Design: “A special research mode where the explorative, generative and innovative aspects of design are engaged and aligned in a systematic research inquiry” (Sevaldson 2010, p.11). Research by design may take different forms. Much of our project was carried out in Rome, surrounded by real life, drawing on participatory design. This is the mode Sevaldson called “the Field” (p. 19). To aid us in the systematic enquiry, we used Gunnar Liestøl’s synthetic-analytic method (Liestøl 2003, 2006). This is a proposition for using humanistic analytical theory not only to analyze texts after they are written, but also to compose new texts. From the analysis of earlier works, one lifts techniques and devices, usually called conventions in Liestøl’s vocabulary, and uses them in new contexts. The results can then be analyzed again, not just for evaluation, but also to further theoretical insight into digital composition or communication.

Observation in the field

The application was developed through a series of iterations. We spent four days in Rome to produce the first prototype, and visited a selection of Roman churches to observe the behavior of visitors to the churches. (Being foreigners helped us in this observation, as we were tourists like the others.) Inside the churches, we realized the profound importance a location may have. Baroque churches in Rome are splendid rooms, visually rich, and with an atmosphere that is strongly felt. The majority of visitors walked slowly into the church and paused to take in the total view of the church. Most visitors also sat down on the benches, either before or after a slow stroll around. We
also took note of the different architecture, lighting and mood in the churches, and tried to write down whatever questions formed in our minds.

With the iPhone as a voice recorder, we produced a few prototype audio commentaries in a form borrowed from radio, and went back to the churches to listen the commentaries together with music. A few experiments revealed that a rather slow pace of reading, mirroring the visitors’ solemn pace, and duration of about one and a half minute for the commentaries worked well. What may feel like a natural tempo in a recording studio or a design presentation will often be too fast for the tranquil mood in a Roman church.

After a series of sketches, an interactive demo was created in Flash (Figure 4) and discussed in several research seminars.

Figure 4: One of many prototype sketches (left) and interactive Flash demo (right). The “tab bar” with four different icons at the bottom proved to be too complex, and was removed in later iterations.

Analysis of earlier genres

Following Liestøl’s synthetic-analytic method, we turned to analysis of earlier genres. There is a growing body of projects developing locative sound (Knowlton 2002; Levine 2005; Rueb 2005; Hight 2006; Løvlie 2009). Most of the works we have found tend to transform spaces into something new by adding
new sounds, or place pieces of literature in an area of a city. We have found few projects that, like ours, try to bring the listener back in time, one exception being the "Chopin benches" in Warsaw. 16 benches of polished black marble are placed in the Polish capital. Each bench designates the place of a significant event in the early life of Frederic Chopin, and can play music and commentary from embedded loudspeakers (“Visit Chopin” 2010).

Technical issues are also the focus in much of the literature on museum and city audio guides, but some have also studied interface and multimodal affordances. Already in early experiments such as the HIPS system, audio gave visitors information while they look at the museum collection (Oppermann & Specht 1999). In the Costa Aquarium in Genoa (Bellotti et al. 2002), the Lancaster GUIDE system (Bornträger et al. 2003), and the Marble museum (Ciavaralla & Paternò 2004), researchers found that users mainly look at images to identify places, while audio works well at the destination. Many also often incorporate adaptive technology. The user’s choices are stored, and the system suggests audio clips that might be of interest based on her or his listening history.

Since the early 2000s, audio guides have developed from experimental systems to consumer products. We analyzed several existing sound guides to Rome and other cities, including iPhone apps and audio guides for rent. In a second visit to Rome, we compared audio guides with written accounts, and different kinds of multimodal works (some of this analysis is reported in Fagerjord 2010). We tested all iPhone guides to Rome available at the time: CityGuide Rome by Marco Polo, Guideyou Rome by Comma, Lonely Planet Rome City Guide, Rome (Travelto) by Molinker, Rome audio city guide by Jourist, Rome City Slicker, Rome touristic audio guide by Guide & Tour City Guide, Top 10 Rome by Dorling Kindersley, Visual Rome by Open up, and When in Rome by Frezza. In addition, we used guidebooks from publishers Dorling Kindersley, Politiken, Galimard (Cartoville guides), and Collins (Access guides); and the web guide Schmap (http://www.schmap.com).

The best audio commentaries were written with a focus on the listener's experience in a certain spot. Good writers of locative audio understand the situation at the spot where the listener is, and adapt the text to that very place. Several of the spoken texts in Jourist Verlag’s Rome Audio City Guide are general information about history, and do not direct the listener to take advantage of actually being on the spot by walking or looking in a certain direction, as the clips in the Rome Touristic Audio Guide generally did.
Spoken lectures are easier to understand when the place is specific and uniform. Rome Touristic’s track for Piazza del Campidoglio is one example. Placed in one corner of the square, a listener can easily make out the different details the narrator mentions. Much more difficult is the overview of the Forum Romanum in the same guide, when the narrator tries to explain certain buildings. There is so much to see from the specified viewpoint that it becomes difficult to understand what is described on the soundtrack. Visual text-types, such as an annotated picture or a map, would have made it easier to point out details in a landscape.

The key factor is not necessarily the size of the spot, but the specificity of the theme. The place may very well be big, as long as the audio commentary pertains to the whole, as when standing on top of Colosseum looking down, while the narrator describes the games that once took place there (as in Rome audio city guide and Rome touristic audio guide). The story is tied to the whole building, not a specific point, and the listener is free to let her or his eye wander around. Guides support at least two different experiences: One is to listen to, read or view information about a certain spot; the other is to find your way to that spot. Few guidebooks or iPhone apps were good at both.

The most common problem was inadequate maps. In some apps, markers for places of interest were not labeled (Jourist, Guide & Tour), in others; the labels were difficult to read (Lonely Planet). Even worse were the apps that had a map, but without links between the map and the lists of places of interest (Top10, Travelto). The Travelto app had only long articles summing up entire districts. This is fine for preparing a trip, but meaningless in the street, where you need short articles that can be accessed on the spot. Top10 and Jourist had such articles, but they were difficult to find, as the list items were always sorted in the same way. Lonely Planet, Marco Polo and GuideYou all had the option of filtering by proximity. Pressing a “nearby” button activated the GPS, and then showed places of interests near our location. This worked very well, and we adapted it into our application.

After this analysis, we split the design into two parts, designing the listening experience in the churches, and design of the guiding system that would help users finding the churches. Both parts used results from the analysis of earlier genres.
Design of the listening experience

Our observations of tourists in the churches, and our own experience, told us that most visitors don’t stay too long in one place, so it was sufficient to select one composer for each location. We put much effort in finding music that suited the looks and mood in each church. We chose pieces that are striking in the first 30 seconds. After about 30 seconds, the music was faded to a lower volume, and a commentary was read with the music in the background. When the narration finished the music was faded up again, and continued until the end of the track.

Some of the earliest experiments in audio hypertext (“hyper-radio”) found that the “tell me more” figure worked well, being a very basic figure of dialogue (Britt et al. 1996). We incorporated a very basic version of the same figure: For each church, an introductory track plays back in the “inside screen”, then one or two additional tracks can be accessed. We prototyped several kinds of links and dialog boxes for this, but all proved too difficult to use, so we ended up with using a track list, similar to the lists of album tracks in iTunes, Apple’s music player app.

Movement design: paper prototyping

For the experience of locating the churches, we carried out three rounds of paper prototyping and mockups, and two rounds of testing on the iPhone. We drew simple pencil sketches of different screens and modes, and simulated the workings of the interface by moving from one sketch to another, as described by Carolyn Snyder (2003). We walked for several days in Rome pressing imaginary buttons in a notebook, also inspired by the design process for the Palm pilot, as reported by Bill Moggridge (2007). When the prototype seemed to work, the design specification was handed over to our iPhone programmer. A first version was tested informally with a few friends of the researchers, and the next step at the time of writing will be to develop a fully functional beta, which will be user tested more formally in Rome.

As we have worked through these development stages (the synthesis in Liestøl’s method), we have made several analytical insights on composing locative sound. Below, we will discuss how it is to compose a hand-held hypertext, issues of navigation in a city as well as in an information space, and finally, the experience of a place augmented by music and commentary.
Hand-held hypertext

Any design project will face the limits of the chosen technology, in our case; it is the small screen of the phone, which limited the use of cross-referencing.

Ever since Vannevar Bush (1945) and Ted Nelson (1974/1987), hypertext has been thought of as cross-connecting associative links showing the many-dimensional connections in a topic. In our early sketches, we used a map, a timeline, three different lists (Churches, Composers, and Works), all available from a “tab bar” at the bottom of the screen (see Figure 4).

We saw quite early, however, that design strategies from desktop applications do not necessarily result in effective design for handheld hypertext. A tourist walking in the busy streets of Rome will not be able to negotiate artful links and connections in a hypertext (and will probably not be interested either, Rome being the city it is). If a hand-held hypertext is to be effective, the author must have the courage to prioritize.

Furthermore, there is not much room for links on the screen. A map will almost always be a zoomed-in partial map. Photos will be viewed on at a time, accompanied by only a few lines of text, and you can never show a map and an image at the same time. Links will have to be few, as our fingertips have difficulty in selecting small targets, and if they sit close together, it is easy to press the wrong link. A handheld hypertext will have to be made up of many short nodes with few links, and in practice, many lists. Still, lists of links have a long tradition in hypertext research. George Landow’s early hypertexts contained what he called “overviews” which were essentially lists of links, leading to topical nodes (Landow 1991, 2006).

Showing users the way

When doing research by design in the field, user needs come first. Our analysis of existing guide application showed that users need a well thought out system to find their way in Rome’s busy streets. Before a user can sit down and listen, he or she must find the church. Most of our design iterations were refinements to the list of churches and the map. Our early sketches had five different views, similar to the Lancaster GUIDE and the Marble museum demo (Bornträger et al. 2003; Ciavarello & Paternò 2004). During testing, we were able to prune it down to the list and map views (Figure 1), as we increasingly realized that proximity was much more important to users than other ways of sorting the material, such as list of composers or a timeline.
Figure 5: Navigating towards a location.

The sequence of screens is modeled to a set of concentric circles (see Figure 5). In the outer circle, a user will use the GPS and map to find the directions from where s/he is to the church of interest. When s/he can see the church, she or he is moving to the middle circle, a screen that describes the church with words and a photograph. To enter the church is to move into the inner circle, where music is played. The succession of screens is a line that, like a compass needle, always points towards the centre, wherever the user starts.

*In place experience*

A simple, pared-down design was also important for the experience inside the churches. Everyone who has collaborated on this project has at some point suggested including photographs in the application. But as our analysis showed (see section 0 above), images on a phone screen make little or no meaning in a Roman church. The church is so big, so rich, that to look at a tiny screen for a photograph of the same would be ridiculous. Sound, on the other hand, works well. The churches are mostly quiet. A spoken commentary or a music track adds to the visuals, while the user is free to look around and take in the splendid surroundings.

In line with the principles described under 0 above, we kept our own comments general, applicable to the churches' overall style. We endeavored to find music that fit the overall style and mood in each church. Several of Rome's basilicas had active chapel masters, choirs, and even instrumental ensembles for centuries. Initially, we thought that we would offer a menu in these churches, so the user could choose which composer or time period to listen to. In our tests on site we discovered that a menu would take to much
attention away from the site itself. Instead, we chose to give priority to one composer (or characteristic style) in each church.

We preferred to find music that echoed with the most prominent style in the church itself. Music, art, and architecture within a given epoch share a common aesthetic, and that is actually felt on the site. When we tested different kinds of music inside a church, some recordings just felt right, as they were in harmony with the architecture and art, and we hoped to find such nice fits in every locale. Such a lucky match was not always possible, however. Georg Friedric Handel, e.g., is one of the most famous composers who ever worked in Rome, and his oratorio *Dixit Dominus* was premiered there. But it was performed in a small, intimate church, which does not at all resemble Handel's powerful work. Instead of a fit, the music contrasted with the room. Still we kept it, for the historical interest, as we did in a couple of other places.

Roman churches are not period pieces in the way the music is. Every church has art from several centuries, and many are also rebuilt and changed since the music was written. Still, the music and commentary is selected and written with focus on a certain dateable aesthetic. Visitors will notice certain aspects of the church, an effect that heightens the experience of time travel, but still to a time that is mostly imaginary. John Urry has described how tourists look upon historic and other places differently than what locals do (Urry 1990). Tourists, defined as people traveling when not working, spend more time studying details, than non-tourists (locals, doing their daily business). Urry calls this the “tourist gaze”. Our application is prescribing a tourist gaze that is quite different from how members of the church gaze when attending a service, or visiting the church to pray. We are transforming a public space with our app. This has some similarities to effect of location-based games Adriana de Souza e Silva (2006) described as hybrid reality, although our project clearly falls out of her definition.

Our experiment is similar in spirit to Augmented Reality (AR), but does not have the real-time images that are the hallmark of AR (Milgram & Colquhoun 1999; Höllerer 2004). It could also be described with Manovich's (2003) term “augmented space”. The effect is quite special. Our church music in Rome app creates an effect of a place that is not entirely real or present-day, but not entirely fictional or historical. The visitor is invited to view the present state of a preserved historical building, layered with visible (and impressive) monuments from several centuries, while listening to a modern-day performance of music written while parts of the interior were made. Both the
place and the music are modern versions of old art, lovingly preserved, but also still used in ways that are felt to be relevant today. Combined, music and place makes the visitor focus on the aesthetic of a past time. The sounds from the present are muted or shut out, and the visitor is encouraged to imagine the same place in a different time.

Conclusion

What are the affordances of sound in locative media? How can one design an interesting and enjoyable location-based learning experience in audio? Is multimodal composition for a specific place different from less specific kinds of authoring?

The early results from this experiment indicate that situated sound can function well for culture communication and learning. Our tests confirmed earlier research that found it effective to address the ears while the eyes take in the surroundings (Oppermann & Specht 1999; Bellotti et al. 2002; Bornträger et al. 2003; Ciavarall & Paternó 2004). However, we repeatedly found that location-aware texts require location-aware authors to be effective. Our initial analysis of existing sound applications and guides showed a marked difference between different texts. Location-aware authors understand what we found when analyzing audio guides, that those texts that have references to what the user can see on the site are far better than the others.

Furthermore, one must understand how users are likely to behave and move in that space, based on its look, sound, atmosphere, and day-to-day use. A hand-held text is not likely to influence much on this. This means that text, visuals, and sound must be authored so it fits with the user's likely behavior, whether it is in the busy streets of a large city, or in the solemn atmosphere of an old basilica. In the words of Belotti, et al. (2002), “added-value information ... should be synergetic with a visitor’s direct ... experience” (p. 35). Further research is needed to establish guidelines for writers of situated sound, what might be called a rhetoric for situated texts.

Apple’s iPhone User Interface Guidelines stress that mobile applications should be simple and have few features (Apple 2008, pp. 36-39). This has also been found in earlier research on location-aware hypermedia systems (Ferris et al. 2010; Häkkilä & Isomursu 2005; Nielsen 2010). An app for situated sound must be simpler still. We started with an interface similar to earlier systems, and found that it was better if we reduced the number of on-screen
controls. What matters most is the spoken commentary, images and interface elements should be kept at an absolute minimum. Still, for some genres, such as education, there might be a need for situated texts that are longer and more complex. Finding formats and interface elements that can support longer texts is a future goal. The iPhone’s small screen is also a limiting factor. Newer tablet computers, such as Apple’s iPad, may enable simple interfaces for more complex hypermedia.

Acknowledgements

Many thanks to Synnøve Kleive (Bouvet ASA) and Anthony Perritano (Intermedia, University of Oslo) who participated on this project. Thanks also to the Inventio project (University of Oslo) for partial funding, and for discussions and suggestions. Inventio is funded by the Norwegian Research Council.

References


