Enhancing Musical Experience through Proximal Interaction

Atau Tanaka, Antoine Authier, Guillaume Valadon, Laurent Violette
Sony Computer Science Laboratory Paris
tanaka@csl.sony.fr

Abstract

The availability of audio content on networks has indelibly changed access to music. We extend this dynamic to wireless ad-hoc networks and present the use of dual infrastructures to extend modes of personal and interpersonal music listening. We identify two distinct forms proximal interaction, sequential and parallel, and demonstrate a musical application for each. Mobile handheld terminals become prototypical future personal music players. They situate the listener in a wireless network space, helping him to define a personal musical sphere with respect to his architectural and social surroundings.

1. Introduction

Networks have changed the way people listen to music. Peer-to-peer file sharing have exploded traditional music distribution channels [1]. Personal music players based on mass storage devices allow one to go jogging with his whole album collection. The first phenomenon has created enormous problems for the record industry. The second has created a boon in the amount of music a listener has at his fingertips. The true impact of network technologies on music, however, goes beyond the immediate commercial impact, or short term consumer satisfaction. Ultimately these technological evolutions lead to a fundamental shift in our relationship with music and with sound as a medium. Furthermore, these changes in perceptual audio interaction have the potential to facilitate human social interaction. This paper looks at the influence that wireless networks can have on music listening habits. We take the social impact not only as a point of departure but as a technology driver in human-centered design.

The peer-to-peer revolution has shifted the focus of music propriety from ownership to access [2]. This creates a need for new techniques for information navigation, addressed by research in meta-data and semantic data retrieval [3]. Meanwhile, the proliferation of wireless protocols and devices have created a need for intelligent routing schemes [4]. The interaction of mobile devices with network based data have led to concepts of ticket based access [5]. This has led to the application of research in human computer interaction (HCI) to the domains of networked devices and media content systems. The vulnerability of wireless networks and the relative complexity of parametrizing configurations have created a need for intuitive user interfaces that allow secure interaction. The present paper expands on the notion of *proximal interaction* to propose new modes of personal music listening.

Proximal interaction is defined as the use of a near-field communications channel to facilitate subsequent interaction on a wireless network [6]. The near-field channel can be characterized by short transmission range, direct contact, or line-of-sight. The proximal interaction model claims that wireless communication can be made more intuitive and secure when initiated by a near-field channel.

Here we present two applications of proximal interaction to music listening. The first uses infrared data (IrDA) [7] to bootstrap ad-hoc Wi-Fi (IEEE 802.11b) [8] networks allowing peer-to-peer wireless audio streaming. The second demonstrates the simultaneous use of Bluetooth [9] and 802.11b to create a location sensitive home audio system. The first uses social dynamics to create private wireless networks independent of the existence or absence of public hotspots. The second example establishes a *personal musical sphere* that is sensitive to user's movements from room to room in a home.

In the following sections, we first describe the two applications. We then show how they represent two distinct modes of proximal interaction. In the conclusion, we look at how this goes beyond just ways of listening to music, to how this enhances the user's relationship and interaction with his architectural and social environments.

2. Ad-hoc Music Sharing

The first example is the use of proximal interaction in its classical sense, whereby a near-field channel initiates communications on a wireless network. In this case, the technique is used to facilitate music sharing across mobile devices. The system allows spontaneous bootstrapping of private wireless networks by intuitive actions, enabling a user to share the music on his device via streaming to a friend's device (Fig 1).

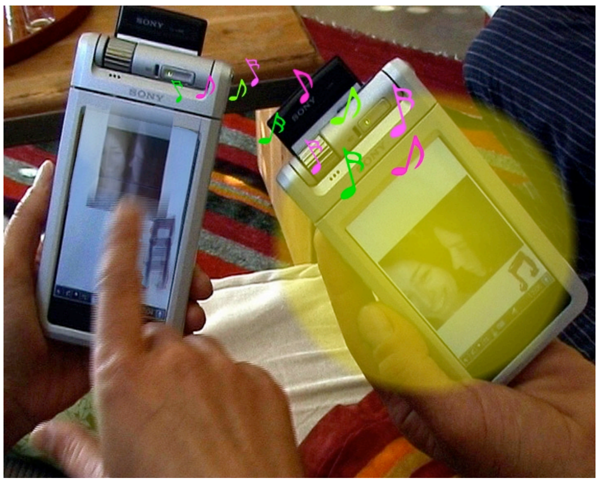

Fig. 1. Spontaneous Ad-hoc Music Stream Sharing

The system uses IrDA as the near-field channel to bootstrap an Ad-hoc 802.11b infrastructure. At the beginning of the transaction, two users each have similar portable music players that are IrDA and Wi-Fi enabled. One of the users is listening to a piece of music and would like to allow his friend to listen in on his own device. They would thus like to establish an instant network independent of their environment, e.g. to create a Wi-Fi network without being dependent on the presence of a public *hotspot*. The goal is to enable these two users to create an 802.11b network in *ad-hoc mode* without having to explicitly parametrize their devices. The act of setting up the network should be intuitive and should follow the social rhythm of two acquaintances meeting, and should not involve distracting tasks such as entering usernames, machine id's, or authentication passwords.

User 1 uses IrDA to transmit (*beam)* session information to User 2. This act indicates User 1's willingness to share the music he is currently listening to. The beamed data contains 802.11b configuration parameters such as SSID network name, IP address, and WEP encryption key. IrDA as a short range, line-of-sight infrastructure, is assumed to be secure. This initial *direct contact* is used to bootstrap a second, wider range network. The security of the initial transaction assures that the main infrastructure created is a trusted one.

User 2 then sends a request to the device address of User 1, indicating his desire to join User 1's network and listen to the music he is listening to. This takes the form of a simple HTTP request invoking a web server running on User 1's device to start streaming the current musical selection simultaneously to User 2's device. The users, with only two button presses after a serendipitous meeting, are successfully listening to the same piece of music each on their own personal device. All subsequent communications, which could include playlist exchange, associated image data, and further audio streaming, take place on the 802.11b channel. As it is a network that has been established in confidence (via the initial near-field exchange), this communications is secure to eavesdropping by other nearby Wi-Fi devices. At the same time, the users may choose to expand their network and allow other friends to join, via the same proximal interaction.

3. Location Sensitive Audio Redirection

The second example is a network architecture for a location sensitive distributed home audio system. The system uses proximal interaction to track user displacement from room to room and redirects a musical flow in such a way that the music "follows" the user. The user has a handheld device that serves as interface to the system. Audio, streaming from a network source, is automatically delivered to the sound system in the room where the user is.

The system consists of several components deployed within the environment of the home (Fig. 2):

- Handheld device capable of simultaneous use of Bluetooth and Wi-Fi
- Bluetooth access points in each room
- Network enabled audio speaker systems
- Music content source streaming MP3
- Network relay, implementing Multicast [10] and TCP relay [11]

These components interoperate over two parallel infrastructures, using Bluetooth as the *near-field* channel and 802.11b as a *pervasive* infrastructure. The near-field channel is used for localization whereas the pervasive channel is used for interaction and content delivery.

The handheld serves two purposes, first as a location signal transmitter and as a graphical user interface (GUI). The Bluetooth interface on the handheld is *discoverable*: as the user moves from room to room with the device, each probe detects its appearance or disappearance from its coverage field. The probes regularly send a list of discovered devices over UDP to a multicast node, a virtual network address that serves as a location reporting point. A server polls and reconciles incoming location reports and deduces the room where the handheld, and therefore user, are.

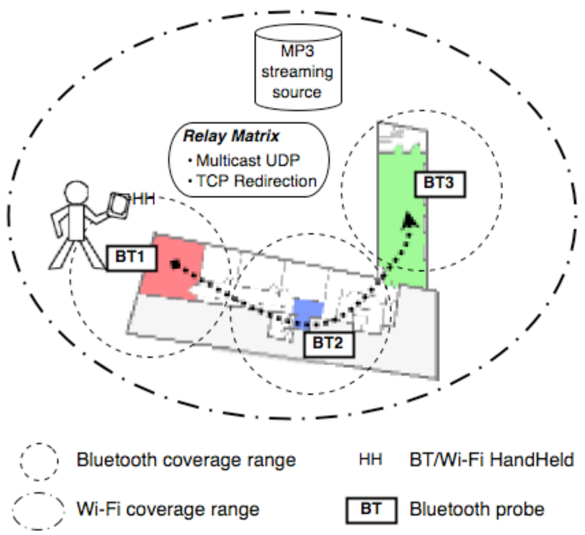

Fig. 2. Localization system components in the context of short-range and pervasive infrastructures of a home

Based on this location information, the server instructs a TCP packet-relay module to dynamically redirect a stream of incoming audio data to the corresponding destination speaker. The incoming audio is the current musical selection, originating from a source as MP3 encoded audio streamed over HTTP. The destinations are HTTP-enabled speaker/amplifier systems constituting the sound system in each room. As the user moves within the home, the music will be re-streamed to the room he has just entered. The music continues in a seamless fashion, picking up in the new room where it left off in the previous room.

While the Bluetooth signal drops off outside each room, 802.11b gives pervasive coverage. This serves to maintain a persistent connection with the handheld to dynamically download new GUI elements, and to capture user action. User actions on the GUI are sent as XML messages over the 802.11b connection to the server. The server parses the incoming messages, interpreting them within the global context, and redispatches the message to the appropriate recipient. The end result for the user is that he has the sensation that the music he is listening to is always with him, and that he has intuitive control over the music via the GUI on the handheld. All data routing and command dispatching are handled by the system without direct intervention on the part of the user.

4. Discussion

While these two examples demonstrate the usefulness of proximal interaction in music listening, they also point out two distinct modes of proximal interaction. The first example uses a near-field channel to configure a pervasive infrastructure. As there is a switch-over from one infrastructure to another, we characterize this mode of proximal interaction as *sequential* (Fig. 3).

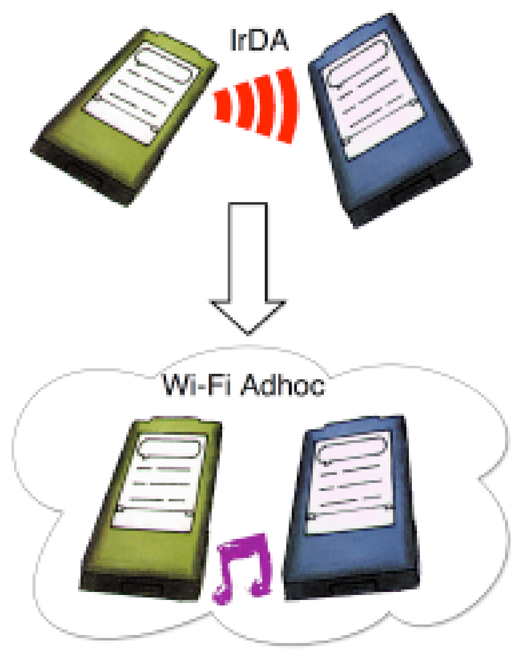

Fig. 3. ***Sequential Proximal Interaction*****: IrDA used to bootstrap an ad-hoc Wi-Fi network**

The second example shows the simultaneous use of the near-field and pervasive infrastructures. This results in an *overlay* structure (Fig. 4) of two independent infrastructures. Functionality is isolated within each layer, i.e. localization is handled by Bluetooth while audio delivery and user interaction take place in the Wi-Fi layer, however information from one layer can influence actions in the other. This is facilitated by system modules that serve as *bridges* between the two layers. In our case, the bridges are the multicast node and the handheld. We call this mode of proximal interaction, where an overlay exists between the near-field and pervasive channels, *parallel proximal interaction*.

These two modes of proximal interaction permit a different kind of use of wireless networks for music listening. The sequential mode allows spontaneous networks to be created regardless of the environment. Interestingly, even if these friends were to meet in a café that *has* a Wi-Fi hotspot, the ad-hoc network that they create still exists as a private interpersonal bubble invisible to the surrounding public infrastructure.

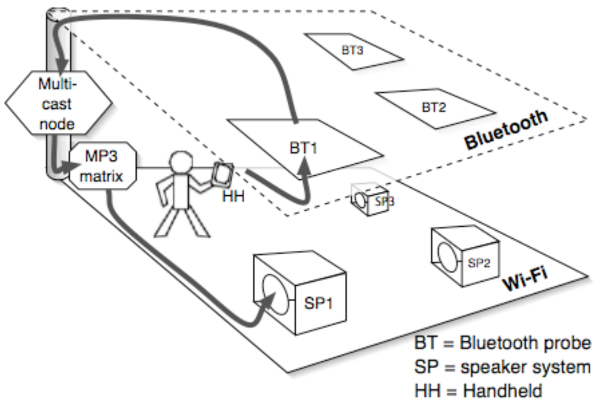

Fig. 4. Overlay view of *parallel proximal interaction*, with data flow passing via bridges

The incremental nature of sequential proximal interaction makes it akin to social customs establishing phases of an acquaintance. Upon first meeting, a handshake and exchange of business cards can be compared to near-field interaction. Once established, each has the information needed to contact their correspondent through means other than direct meeting, be it telephone or postal mail, that is to say, to exploit a pervasive infrastructure. This metaphor can be extended to people presenting third parties. Much in the way that a new community member can introduce himself on the behalf of a known member, our system is able to extend the private network through a series of introductions. We are currently creating simulations to follow chains of trust, where public-key exchange among devices takes place based on this social introduction mechanism.

In the first example of stream sharing, the listener has the possibility to *socially* extend his personal music sphere. In the second example, the user has the sensation of *architecturally* extending his musical sphere. From the user point of view, the music becomes a ubiquitous medium – it appears in every room he occupies. From the architectural point of view, it is in fact a focused personal sphere surrounding the user. This sensitive redistribution of sound leaves acoustical space available to be occupied by the musical spheres of other listeners. We have begun to look at the *collision* effects when two peoples' musical spheres intersect. Collision resolution is not just a question of giving dominance to one music over the other. Satisfying solutions call for sophisticated acoustic and signal reconciliation techniques.

The systems presented here show the pertinence of proximal interaction in creating personal music systems. The net effect is the use of principles of ubiquitous and calm computing [12] applied to enhance the act of music listening. Music is a good test case for these techniques for its nonvisual and pervasive qualities as an acoustic medium, and for its social aspects as content commodity. The systems illustrated here demonstrate the use of wireless networks not just as a convenience factor in music access, but as infrastructures on which listeners can tailor their personalized musical environments.

5. References

[1] Alderman, John, *Sonic Boom: Napster, P2P and the Battle for the Future of Music*, Fourth Estate, London, 2002.

[2] Rifkin, Jeremy, *The Age of Access: The New Culture of Hypercapitalism, Where All of Life Is a Paid-For Experience,* J. P. Tarcher, Los Angeles, 2000.

[3] La Burthe, A., Pachet, F. and Aucouturier, J. J., "Editorial Metadata in the Cuidado Music Browser: Between Universalism and Autism", *Proceedings of the WedelMusic Conference*, Leeds, 2003.

[4] Costa, L., Fdida, S., and Duarte, O., "Hop by Hop Multicast Routing Protocol", *ACM SIGCOMM '01*, San Diego, 2001.

[5] Cho, Kenjiro, "Ticket-based Device-to-device Trust Management through Tactile-Space Networking", *Unpublished manuscript*, 2003.

[6] Jun Rekimoto, Yuji Ayatsuka, Michimune Kohno, Hauro Oba, "Proximal Interactions: A Direct Manipulation Technique for Wireless Networking", *Proc. INTERACT 2003 Ninth IFIP TC13 International Conference on Human-Computer Interaction* Zurich, 2003.

[7] Infrared data association home page. http://www.irda.org

[9] IEEE Wireless LAN Working Group http://grouper.ieee.org/groups/802/11/

[8] Bluetooth SIG. Bluetooth sig home page. http://www.bluetooth.org

[10] Meyer, D. *Administratively Scoped IP Multicast Network Working Request for Comments: 2365,* http://www.ietf.org/rfc/rfc2365.txt 1998.

[11] Icecast, http://www.icecast.org

[12] Weiser, Mark, "The computer for the twenty-first century", *Scientific American*, pages 94–104, September 1991.