Autonomous Interactive Intermediaries: Social Intelligence for Mobile Communication Agents

Stefan Marti
Speech Interface Group, MIT Media Lab
20 Ames Street
Cambridge, MA 02139
+1 617 253-8026
stefanm@media.mit.edu

ABSTRACT

This paper describes ongoing thesis work about embedding elements of human style social intelligence into an agent that controls a user's mobile communication devices in order to make these devices more socially acceptable to both user and co-located people. Today's mobile communication devices are passive tools that serve as mere communication portals. I propose an *Autonomous Interactive Intermediary* (AII) that can act as an actively mediating party between user, caller, as well as co-located people. The AII is a dual conversational agent that can converse simultaneously with both user and caller. It has a physical embodiment that uses socially intuitive non-verbal cues to express itself in a subtle yet public way. The AII can rely on several methods for harvesting *residual social intelligence* from human and other sources to behave in a socially appropriate way.

Keywords

Computer mediated communication, mobile communication, intermediary, robotic embodiment, conversational agent, human style non-verbal cues, social intelligence.

1. INTRODUCTION

It almost seems like we are 'genetically incapable' of ignoring a ringing telephone. The desire to answer a call sometimes seems to outweigh the importance of maintaining the flow of face-to-face conversation [3]. Even with voice mail and caller ID, co-located people are being ignored in favor of callers [2]. Handling cellphone calls appropriately has become an important social skill. However, the social impact of mobile communication on co-located people is still not a very well researched area.

Our mobile communication devices rarely assist us in managing the transitions between face-to-face and remote conversations. Instead, if left on their own, they interrupt us at inappropriate times such as in a theater or during an important meeting. Although some devices allow us to manually set profiles for certain situations or caller groups, and associate audible and/or tactile alerts with these profiles, we still can regulate our accessibility essentially only in a binary way: device alerting on or off. This results in an unsatisfactory trade-off between not being disturbed and possibly missing a call, versus not missing any calls and possibly being disturbed unnecessarily.

Unfortunately, such unacceptable interruptions do occur often. There is a class of situations when we do not want to get interrupted at all (e.g., important meeting) and therefore will switch off our devices, and another class of situations when we welcome any interruption (e.g., sitting alone in a bar). However, these two classes are just at the extremes of the dimension "openness to interruption." A bigger part of our daily lives falls into the gray area in between these extremes, where it is not so clear if our cellphones should be switched on or off. For example, we usually do not want to get interrupted when we are asleep at night or during our family dinner, unless it is extremely urgent.

Our mobile communication devices may eventually try to tackle these problems by assuming a proactive role. However, it is part of our human nature that we may insist on interacting with both human and other agents in a socially appropriate way, and expect the same social behavior from these agents.

2. RATIONALE

In my thesis work, I propose an *Autonomous Interactive Intermediary* (AII) that plays the role of an actively mediating party between a user and his/her callers, as well as the co-located people. It can fall back on several sources of *residual social intelligence*. The AII is not inherently intelligent as a 'stand-alone A.I.', but harvests 'leftover' social intelligence from nearby sources, both human and artificial. These sources of social intelligence, or modules, can be used separately or together with other modules, depending on their availability, based on the idea that several complementary approaches for intelligence should be used in parallel to create a less brittle system.

One of these sources of social intelligence is the 'local others.' Allowing all participants of a face-to-face conversation to 'veto' unobtrusively and anonymously to an upcoming interruption by a cellphone might increase the social acceptance of mobile devices, shifting the responsibility for an interruption to the group.

Social intelligence manifests itself not only through *reasoning* with social intelligence, but also via *behaving* with social intelligence. Therefore, using non-verbal cues of a robotic user interface—the physical embodiment of the AII—to interrupt and alert a user may be an intuitive way to generate subtle but still public alerts [1] for mobile communication devices.

CSCW'04*, November 6–10, 2004, Chicago, Illinois, USA.*

3. SYSTEM

3.1 Physical Embodiment

The AII has a physical embodiment (animatronics device) such as a parrot on the shoulder, or a small stuffed animal in the user's chest pocket or hand. Strong and intuitive social non-verbal cues like opening eyes and turning head are used for alerting and attracting attention, instead of ringing or vibration. This subtle yet public alerting is important to make the interruption socially more acceptable for co-located people. The embodiment also serves as a natural focal point of attention when the user listens and talks to the AII. Figure 1 shows one of the current embodiments, a wireless computer remote controlled bunny with four degrees of freedom (neck, eyes) and built-in two-way audio channel.

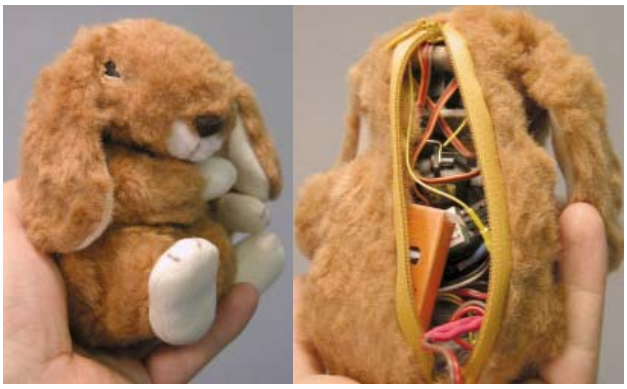

Figure 1: Current embodiment of AII

3.2 Conversational Agent

In addition to its embodiment, the AII is also a dual conversational agent capable of being involved in two concurrent conversations, one with the user, whispering in his/her ear and listening, and one with the caller on the phone—at the same time, mediating between them. The conversational script changes dynamically depending on the identity of the caller, the conversational status of the user, and the content of the call. When interacting with the user through the robotic user interface, the AII only interrupts when socially appropriate, waiting politely for pauses in the conversation of the user with local others. Mediating between user and caller, it is able to break down a synchronous phone call into a series of asynchronous voice instant messages.

3.3 System Behavior

Upon an incoming call, the AII assigns a preliminary priority level (based on caller ID, communication history, etc.), determines local social context based on detection of conversational groupings (*conversation finder*), alerts and allows input from co-located conversation members if they prefer not to be interrupted (*social polling*), and consults the current "place" in which the recipient is located for hints in the form of patterns of prior interactions in these locations (*room memory*). In parallel, it engages the caller in a conversation, asking about the nature of the call and offering to take a quick voice instant message. It then re-classifies the relevance of the call based on the caller's answers to questions as transcribed by speech recognition (*issue detection*). Also in parallel, the physical embodiment alerts the user: the animatronics wakes up, looks around, seeks eye contact with the user, etc. These composite behaviors are generated dynamically from a library of atomic behaviors, depending on the interaction with the user as well as the ongoing conversation of the AII with the caller. The AII then can tell the user about the ongoing call, playing back voice instant messages, and allowing the user to leave an answer for the caller on hold.

3.4 Modules of Residual Social Intelligence

3.4.1 Conversation Finder

This module provides the AII with the current conversational status of the user. This is done via a self-organizing network of small body-worn sensor nodes that communicate over low-range RF when their wearers are speaking. It is used to detect the size of a conversational group, if the user is mainly talking or listening, etc. Protocol and hardware have been developed on breadboards, and badges (38x33x15mm), to be worn close to the user's neck, are at their final stage of implementation.

3.4.2 Social Polling

This module allows the AII to subtly query all participants of a co-located conversation about the appropriateness of an interruption. All participants, detected by the conversation finder module, will be able to 'veto' to an upcoming interruption by a mobile device unobtrusively and anonymously via a wirelessly actuated finger ring that vibrates slightly (work in progress). The AII then considers these vetoes when deciding to interrupt or not.

3.4.3 Room Memory

This module allows the AII to gather information on how mobile communication devices are used (turned off, on, etc.) in the current room sized area by querying a wireless local sub-agent that sums up communication events and broadcast them to interested AIIs. Hardware, software, and ad-hoc networking protocol are similar to the Conversation Finder infrastructure.

3.4.4 Issue Detection

This module allows the AII to assess the relevance of an incoming phone call by conducting fuzzy inferences from recognized words of its conversation with the caller, with what it thinks is currently 'on the mind of the user.' Information for the latter is harvested continuously from the user's ToDo list, recently made web searches, and sent email messages, together with more long-term information mined from the user's personal web pages. These harvesting processes are in operation.

4. REFERENCES

[1] Hansson, R., Ljungstrand, P., and Redström, J. (2001). Subtle and Public Notification Cues for Mobile Devices. *Proceedings of UbiComp 2001*.

[2] Laufer, P. (1999). Wireless Etiquette: *A Guide to the Changing World of Instant Communication.* Cedar Knolls, NJ: Omnipoint Communications.

[3] Plant, S. (2000). *On the Mobile: The effect of mobile telephones on social and individual life.*