Autonomous Interactive Intermediaries:
Social Intelligence for Mobile Communication Agents

by

Stefan Johannes Walter Marti

M.S., Special Psychology, Philosophy, and Computer Science,
University of Bern, Switzerland, 1993
M.S., Media Arts and Sciences,
Massachusetts Institute of Technology, 1999

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Media Arts and Sciences at the
Massachusetts Institute of Technology

June 2005

Author Stefan J. W. Marti
Program in Media Arts and Sciences
May 6, 2005

Certified by Christopher M. Schmandt
Principal Research Scientist
MIT, Media Laboratory
Thesis Supervisor

Accepted by Andrew B. Lippman
Chair, Departmental Committee on Graduate Students
Program in Media Arts and Sciences

Autonomous Interactive Intermediaries:
Social Intelligence for Mobile Communication Agents

by

Stefan Johannes Walter Marti

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
on May 6, 2005,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Media Arts and Sciences
at the Massachusetts Institute of Technology

Abstract

Today's cellphones are passive communication portals. They are neither aware of our conversational settings, nor of the relationship between caller and callee, and often interrupt us at inappropriate times. This thesis is about adding elements of human style social intelligence to our mobile communication devices in order to make them more socially acceptable to both user and local others. I suggest the concept of an Autonomous Interactive Intermediary that assumes the role of an actively mediating party between caller, callee, and co-located people.

In order to behave in a socially appropriate way, the Intermediary interrupts with non-verbal cues and attempts to harvest 'residual social intelligence' from the calling party, the called person, the people close by, and its current location.

For example, the Intermediary obtains the user's conversational status from a decentralized network of autonomous body-worn sensor nodes. These nodes detect conversational groupings in real time, and provide the Intermediary with the user's conversation size and talk-to-listen ratio.

The Intermediary can 'poll' all participants of a face-to-face conversation about the appropriateness of a possible interruption by slightly vibrating their wirelessly actuated finger rings. Although the alerted people do not know if it is their own cellphone that is about to interrupt, each of them can veto the interruption anonymously by touching his/her ring. If no one vetoes, the Intermediary may interrupt. A user study showed significantly more vetoes during a collaborative group-focused setting than during a less group oriented setting.

The Intermediary is implemented as a both a conversational agent and an animatronic device. The animatronics is a small wireless robotic stuffed animal in the form of a squirrel, bunny, or parrot. The purpose of the embodiment is to employ intuitive non-verbal cues such as gaze and gestures to attract attention, instead of ringing or vibration.

Evidence suggests that such subtle yet public alerting by animatronics evokes significantly different reactions than ordinary telephones and are seen as less invasive by others present when we receive phone calls.

The Intermediary is also a dual conversational agent that can whisper and listen to the user, and converse with a caller, mediating between them in real time. The Intermediary modifies its conversational script depending on caller identity, caller and user choices, and the conversational status of the user. It interrupts and communicates with the user when it is socially appropriate, and may break down a synchronous phone call into chunks of voice instant messages.

Thesis Supervisor: Christopher Schmandt
Title: Principal Research Scientist

Doctoral Dissertation Committee

Thesis Advisor	Christopher Schmandt Principal Research Scientist MIT, Media Laboratory
Thesis Reader	Cynthia Breazeal Assistant Professor of Media Arts and Sciences LG Career Development Professor of Media Arts and Sciences MIT, Program in Media Arts and Sciences
Thesis Reader	Henry Lieberman Research Scientist MIT, Media Laboratory

Dedication

For Kimiko

Acknowledgements

First of all, I would like to thank my thesis advisor Chris Schmandt for admitting me to the Media Lab (although it took two tries!) and giving me the incredible opportunity to spend time at this extremely inspiring environment for such a long time. Although I really wanted to build autonomous micro helicopters back in 1997 when I arrived at the lab, I think what we have done instead is almost as cool as levitating devices... Perhaps later I'll combine all the things I have learned about speech interfaces with flying robots!

My thesis committee helped me to keep it real: both Henry Lieberman and Cynthia Breazeal gave me extremely valuable feedback.

Everybody knows that the Media Lab would cease to exist and implode without the help of our genius undergraduate collaborators, the MIT UROPs. I had the pleasure to work with Matt Hoffman (we really should write a paper about the dual finite state machine sometime), Quinn Mahoney (your PCB design eventually worked out perfectly!), Mark Newman (you did an incredible job with the Yano head—you can be really proud of yourself!)

The primary peer group at the Media Lab is your research group, and I would like to thank the Speech Group members, past and present, for their support: Jae-Woo Chung (I still can tell you all about electronics and sensor stuff, if you want to), Rachel Kern (you will inherit the bunny, looks like), Vidya Lakshmipathy (the two years we shared an office were an incredible time—let's do that again! And thanks for accepting Nena as a third office mate), Natalia Marmasse (we really should have written this 'manual', you know what I mean...), Sean Wheeler (I am so jealous of your Japanese skills), Nitin Sawhney (that was an inspiring time, when I became your office mate! You were my big brother—although I was older…), Gerardo Vallejo (hey, you are really good at electronics), Keith Emnett (I remember when we were looking for an apartment for almost a summer, and then found tiny Winter street, super close to Star Market), Zeynep Inanoglu (girl, you can dance!), Assaf Feldman (keep it cool, man), Sunil Vemuri (speech group honorary member—what else is there to say), David Spectre (helicopter piloting skills come in handy when controlling an animatronic parrot, right?) Tracy Daniels (I am glad you like Nena so much!)

My friends here: Push Singh (you were able to pick up Nena, congrats, there's not a lot of people who can do that), Barbara Barry (did we start the whole reading group thing back then?), Tim Bickmore (you're the only real circus guy I know), Jennifer Smith (you guys inherited our Winter street place, but then got this super cool summer resort house—what an upgrade), Hannes Vilhjálmsson and Deepa (I remember very clearly when I first visited the lab in 1996, you did not wear shoes! I was impressed), Vadim Gerasimov (it is incredible how much you know about electronics, hoarder boards, transceiver—and you were always around to help me out when I got stuck), Deva Seetharam (we should have given the Weather Tank talk, and published a paper about it), Ari Benbasat (thanks for the Alias DVDs and all your advice on electronics—you are a

walking electronics encyclopedia), Arnaud Pilpre (ahh, monsieur, comment ça va? Your chaponais is amazing!), Jacky Mallet (and your Japanese is better than mine as well! Am I the only one here who doesn't speak it?), Paul Nemirovsky (we should go to parties more often, man), Yuri Ivanov (you are the funniest guy I know), Pascal Chesnais (Canard is down again!! Do you have the key to the roof? ;-), Sky Horton (one word: Moller Skycar—who else knows about this?), Hiroko Yamakido (I wish we would live closer, so we could hang out much more often! That would be fun), Cati Vaucelle (ahh, La Brazilienne… ;-) You have more French charme than anybody I know!), Diana Young (Nena remembers you so well from when you were sitting her! And she liked your rabbit!), Ali Mazalek (hey, is your Bladerunner heli still flying?), Brad Rhodes (you are the coolest general's exam committee member I can think of), Dan Overholt (you don't realize how much you helped me when you showed me how to hack the buddy plug of the R/C transmitter back in 1998—you are so cool!), Gert-Jan Zwart (I will never forget you—remember the conversation we had about hollow diamond vacuum micro spheres, a solid material lighter than air? That kind of stuff triggered my obsession with levitating devices. I will never understand why you had to leave this earth so early. See you on the other side!)

A very important peer group I had was the Media Lab Student Committee: I thought it was kind of cool what we did, being self selected, without having any 'real' powers—but we have come a long way! In particular: Joanie Morris DiMicco (you were an amazing model for the Nena fashion shoot!), Cory Kidd (you are so driven—you could be a one-man-studcom, easily! Thanks for driving the rest of us), Peter Gorniak (I really wonder why we spoke English 98% of the time—for the first half year I even thought you were native Canadian :-), Ben Dalton (whose turn is it with refilling the vending machine, again?), Aisling Kelliher (you were by far the funniest of us all), Andrea Lockerd Tomaz, Edison Thomaz (too bad you didn't stay longer at the lab, Studcom would have profited from you), Lis Sylvan (you guys had some awesome Halloween parties, I will never forget these! Say hi to Misha and Jeremy), Michelle Hlubinka (I hope that we will be able to join you soon in San Francisco!)

Then there are my Swiss friends who are still my friends, even after 8 years in the States: the Diemers (Lily, you are absolutely and by far the most fashionable kid I know! Well, I am sure Manu had some influence there… You know, sometime in the future, you guys will visit us in the States! BTW, Ralphman, are you aware that we have exchanged 8493 emails since I got here? True!), Markus Spielmann (you are the coolest guy ever. Do you remember when we were 15 and we built small R/C ships and tested them under the Aare bridge in Obergösgen? You sank one of my servos! Ten years later, you had Noiseannoys—I had so much fun I in this studio; I think you are my longest standing friend ever—it is amazing), Sandra & Urs Gallauer (thanks for visiting me in the States during my first year! I will always feel close to you guys, no matter how far we are), David Jäggi (I know I will never be as good as a biker as you are, but the accident I had in Southern France was just because I was tired!), Reto Häfliger (I think you are the guy I played in most bands with: remember Phlox, Team Time, etc? I will never forget those times), Jasmin Schmid (you need a record contract to get big!), Thomi Müller (remember Kanti Olten? Miscast? These were so friggin' cool times—as well as all the following times we spent playing music. That’s the worst part about me in the States: not being able to play in my old bands), Urs Friedli (you guys have come a long way since Nostradamus!), Myriam Frey (I wonder what would have happened if I told you 15 years ago that I had a crush on you), Brigitte Sterchi (let’s talk sometime about what happened a few years ago), Dani Weber (we were by FAR the best roadie team EVER!), Elli Luha a.k.a. Kaza (you said several times that I was the one that got away—I think *you* were the one that got away ;-), Snezana Buzarov (call me sometime—I have to catch up with you. How’s Lilly?), Dark Sign, Coroner, Blue 46—being able to work with you, on tour and in studio, and more importantly hang out with you made me the hippest guy ever!

The Media Lab is one of the coolest place on earth, and it consists of people, too: Walter Bender (do you remember the NiF meeting we had in September 1997, when I first suggested the flying micro helicopter? Yeah, that's when it all started), Hiroshi Ishii (you have been so good with Kimiko and me, thank you many many times!), Marvin Minsky (although you did not end up on my thesis committee, I still got influenced a lot by your ideas), Deb Roy (ahh, we really should have built more small robots back then…), Deb Wiedner (thanks for taking care of Nena so many times—she absolutely loved being with you and your family), Polly, Lisa Lieberson, Linda Peterson, Pat Solakoff, Paula Aguilera (I will return the Umatic!), Matt Traggert (for coaching me with my pirate acting), Necsys (what would we do without you guys?), Nicholas Negroponte (do you realize that you were the reason I got here when you sent me an email reply to my request for 'brochures', back in 1995? I bet you don't remember that. I do.)

Of course my family in Switzerland: My siblings Luki and Chrige—only when you are far away you realize what family really means. My parents Helene and Johannes who supported me all these years in a tremendous way—you are the best parents, ever. I mean it, and I love you.

And my family here: Our dog Nena, and absolutely most of all: **Kimiko**! I could not have done it without you, seriously. You are the best. I love you, and I will always love you.

Table of Contents

Table of Figures

1. Summary of contributions

Embedding elements of human style social intelligence into the agent that controls a user's mobile communication devices, her "Intermediary," will make her devices more socially acceptable, less annoying, and more useful to the user and to the people around her.

In the following I will describe the main contributions of this thesis:

1. The Intermediary can fall back on several sources of 'residual social intelligence.' The agent is not inherently intelligent as a stand-alone artificial intelligence, but harvests 'leftover social intelligence' from close by 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 or overlapping approaches for intelligence should be used in parallel.

2. One of these sources is people: caller, callee, and co-located people contribute to and influence the Intermediary's actions, either through spoken language or tactile input. All participants of a face-to-face conversation can 'veto' an upcoming interruption by a mobile device unobtrusively and anonymously by touching their actuated finger rings. This novel 'social polling' of the Intermediary's immediate environment increases the social acceptance of mobile communication.

3. Social intelligence manifests itself not only through *reasoning* with social intelligence, but also via *behaving* with social intelligence. Using non-verbal signals of a robotic user interface—the embodiment of the Intermediary—to interrupt and alert a group of users is an intuitive way to generate subtle-but-public alerts for mobile communication devices, and is perceived as less intrusive than traditional phone interruptions.

4. An Intermediary who can be involved in two concurrent conversations, one with the user, and one with the caller—at the same time, mediating between them, being able to break down a synchronous phone call into asynchronous pieces (chunks of voice instant messages)—allows the user to reduce the time spent on the phone and increase the time spent on face-to-face interactions.

2. Introduction

> “Most of us are genetically incapable of ignoring a ringing telephone. It doesn't seem to matter that modern technology also gives us answering machines and voice mail. Every day, living, breathing human beings get ignored in favor of unknown callers. Even the more recent caller ID feature—which lets us know who is electronically knocking at our virtual door—doesn't stop some of us from giving distant visitors priority over actual, present conversations.” (*Wireless Etiquette*, Peter Laufer, page IX) [105]

> “The ability to handle mobile calls has become an important social skill. A ringing mobile will often take precedence over the social interactions it disrupts: the need or desire to answer a call often outweighs the importance of maintaining the flow of face-to-face conversation. This is why even a silent mobile can make its presence felt as though it were an addition to a social group, and why many people feel that just the knowledge that a call might intervene tends to divert attention from those present at the time. The mobile tends to siphon concentration; for many couples, its presence can be as powerful and distracting as that of a third person.” (*On the Mobile*, Sadie Plant, p 1) [157]

People use mobile communication devices everywhere, all the time. Quite often, they do so even if they are not alone, and therefore, the desire to telecommunicate and to communicate with co-located people simultaneously clashes.

Over a long period of time, the human species has developed efficient ways of regulating and maintaining conversations with co-located people, using a variety of verbal and non-verbal cues, which are well studied by social psychology (e.g., Goffman, 1966) [67]. However, our current mobile telecommunication devices often disrupt these regulatory mechanisms of human conversations (McLuhan, 1964) [132].

Worldwide, the use of mobile telecommunication devices is increasing at a high rate. Ethnographic studies document how this technology is starting to influence all aspects of our lives (e.g., Rheingold, 2002) [167], but especially our social relationships: Sadie Plant’s *On the mobile* [157] is full of beautiful anecdotes that illustrate this influence. The social impact of mobile telecommunication, defined as the impact of mobile communication on the relationships we try to maintain, seems to become very relevant. However, research that aims at understanding this impact is rather scarce (Geser, 2002) [66].

In particular the social impact of mobile communication on co-located people has not often been studied systematically, perhaps with the exception of 'mobile communication etiquette' (e.g., Laufer, 1999 [105], and Ling, 1997 [113]), and a recent sociological study (Geser, 2002) [66]. But as it is intuitive that being alone versus part of a big group of socially active people modifies one’s telecommunication behavior, it is also clear that interacting with remote people via a mobile device can have a strong influence on the social relationships with co-located people.

The simplistic human communication model that expects a sender, a receiver, and a channel is often no longer applicable to situations where a user communicates with a remote party, while remaining part of a group of co-located people. In this situation alone, there are already three participants involved, the co-located person being the third party. This is specifically true for mobile communication, where the caller often does not have any information about the social setting of the callee prior to the call.

As our mobile telecommunication devices mature, they most likely will become another independent entity in this complex multi-party communication scenario. But it is likely that we humans may insist on interacting with all these human and non-human parties in a social manner, and expect them to interact with us similarly (Reeves et al., 1996) [164]. Such expectations will pose a challenge for our mobile communication devices, or rather, for the designers of these devices.

Our mobile device not only lack the capabilities to interact with us in a social manner, but also don’t help us to integrate the two facets of communication, communication with co-located people and telecommunication with remote people using mobile devices. Instead, mobile calls interrupt us at inappropriate times, such as during public performances, during important conversations with our superiors, etc. This is not acceptable for obvious reasons.

Although modern communication devices allow us to set manually profiles for certain situations and caller groups, many still give us only the option to control our accessibility in a binary way—switch the phone off, or leave it on. This results in an unacceptable and frustrating trade-off between not being disturbed and possibly missing a call (as well as upsetting a caller), versus not missing any calls and being possibly unnecessarily disturbed (and upsetting our co-located conversation partners).

Most importantly, the interruption is most annoying not for the callee, but for the bystanders:

> “There is a lack of symmetry in the perceived impact of an interruption. When I have lunch with friends who spend a considerable fraction of our time responding to calls on their cell phones, I consider this a distraction and an interruption. From their point of view, they are still with me, but the calls are essential to their lives and emotions and not at all an interruption. To the person taking the call, the time is filled, with information being conveyed. To me it is empty unfilled time. The lunchtime conversation is now on hold. I have to wait for the interruption to end. How much time does the interruption seem to take? To the person being interrupted, forever. To the person taking the call, just a few seconds. (...) The person engaged in the cell phone conversation feels emotionally satisfied, while the other feels

> ignored and distanced, emotionally upset. (*Emotional design: why we love (or hate) everyday things*, Don Norman, p 153) [150]

Unfortunately, such unacceptable interruptions occur often. Although there is a class of situations where we certainly do not want to receive *any* interruption at all (e.g., important but short meeting with superiors) and therefore will switch off our cellphones, and another class of situations where we will accept *any* interruption by our communication device whatsoever, these two classes describe just the two extremes of the dimension "openness to interruption." For a good part of our daily lives, we are in the gray area in between these extremes, in situations where it is not so clear if our communication device should be switched on or off. For example, when we sleep at night, and during meals with our families, we usually don't want to get interrupted, except if it is very important. Unfortunately, just these two situations—sleeping and eating—can easily account for half of the 24 hours of a day!

It is exactly for those situations—where interruption should only happen if appropriate—that we need communication devices that have (hopefully at least) a small idea about what is going on in our lives, or in other words, have some human style "smarts" built-in.

Computationally advanced mobile devices such as smartphones—handsets resembling standard mobile phones rather than PDAs, yet featuring always on wireless access to IP networks and significant computing power—are not smart in human terms. Human style "smarts" has to be based on intelligent reasoning and intelligent behavior. In the domain of communication, specifically, it comes only with "communication intelligence," or social intelligence. Social intelligence is the ability of people to relate to others, understand them, and interact effectively with them. An advanced hardware platform alone does not make a mobile device smarter, or more useful to the user. In this thesis, I suggest what can be done to make smartphones truly smart.

However, social rules are neither static nor universal, but instead change over time and vary significantly from culture to culture. This thesis does not claim to describe the only possible way in which humans manage interruptions. The approach is rather to assume one point in social space, and demonstrate how technology can assist with managing interruptions by mobile devices.

Figure 1: Apparatus for preventing unwanted cellphone interruptions

Today, the people who interrupt us most often by calling or mobile devices may well be our friends and family, but it is foreseeable that unwanted phone calls and interruptions will increase in the future, creating a similar problem to the flood of unwanted email messages (spam) that all of us receive today. Similar to 'spam filters,' it may be necessary to plan for equivalent measures in the phone domain (and low-tech solutions as in Figure 1 may not suffice).

3. Rationale

3.1. Social Intelligence

My proposed solution hinges on the concept of **social intelligence** (Kihlstrom, 2000) [98]: if our mobile communication devices had human style social intelligence, then they would be more useful to us, and our lives would be easier. Human style social intelligence means that our communication devices would do what we expect them to do, and especially what *not* to do, without being explicitly told.

In order to get to this point, I suggest that a mobile communication device has to be looked at less as a tool or a mere portal to another person, but rather as an "Active Intermediary." Such an entity would act on behalf of its user, but may also interact with the use's remote communication partners, with the people around the user, and her location. In order to do that, it needs to understand the basics of human communication: it needs to have human style social intelligence.

Proposition 1:

Embedding elements of human style social intelligence into the agent that controls a person's communication devices will make these devices more socially acceptable, less annoying, and more useful to the user and the people around the user.

However, for a communication device, acting with social intelligence is a hard task. The device has to know things on a macro-relational level, such as what are the relations of the cellphone owner with the people who communicate with her, what are the goals and desires of the user, what seems to be on her mind at a given point in time that would justify an interruption? It also has to know how people interact on a micro-relational level: when is it appropriate to interrupt the user, given her current social situation? Most importantly, it needs to 'blend in' when the user is part of a group, in a social situation (vs. being alone), using the same subtle signals that humans use to control the interactions between them. Therefore, it needs to be able to express itself with non-verbal behaviors (e.g., eye gaze), mainly to interrupt in a socially acceptable way, but also to express its inner state in an intuitive way to the user and the co-located people. Research has shown that the most appropriate alerting behavior is subtle, but public, making such behavior visible to the people around the user (Hansson et al., 2001) [77]. Towards its user, however, a socially intelligent communication device needs conversational capabilities: with soft-spoken language, possibly whispering in the user's ear, it could summarize an ongoing call or past communication events, much like a secretary that taps on a user's shoulder and whispers in her ear a short summary of something important that just happened, after having waited for her turn to interrupt the user verbally or non-verbally.

3.2. Active Intermediary

Rather than overloading the current notion of a (passive) mobile communication tool with the idea of social intelligence, I suggest unifying all the above mentioned facets of social intelligence into a single, active, independent entity: an *Active Intermediary*. This entity—agent, daemon, persona, angel, etc—would try to take into account the above mentioned aspects of human style social intelligence when acting on behalf of the user and controlling her personal mobile communication infrastructure.

Proposition 2:

Implementing a mobile communication agent into a dedicated entity, an Active Intermediary (as opposed to a passive communication device that serves as mere communication portal), will greatly enhance the possibilities of telecommunication.

3.3. Physical Embodiment

However, social intelligence does not only include *reasoning* with social intelligence, but also *behaving* with social intelligence. Even if a software based Intermediary would be able to reason successfully with human style social intelligence, it would not impact our real world—and the humans living in it—unless it would be able to 'act out' its knowledge. In other words: social intelligence has both artificial intelligence as well as user interface aspects. Embodying a software agent in the real world requires using interface and communication paradigms that we humans are used to.

A common solution to this problem is to embody the agent on a computer screen, e.g. in the form of an animated character. However, such a character doesn't exist in the same physical world as we humans do—rather it appears to be seen through a window. Furthermore, many non-verbal cues depend on three-dimensional space and are lost on flat display screens, therefore diminishing the richness of the communicative exchange (Breazeal et al., 1999) [22].

Embodying the Intermediary into a physical entity places the Intermediary in the same physical world as its user and the co-located people. Studies have shown that a robot is more engaging, credible, informative, and enjoyable to interact with than an animated character because it is a real, physical thing, as opposed to a fictional animated character on a screen (Kidd et al. 2004) [95]. Such a 'robotic user interface' (RUI) (e.g., Bartneck et al. 2001a [7], Bartneck et al. 2001b [8], Sekiguchi et al. 2001 [187]) or 'human interface robot' might enable a very immediate human-machine interface in human terms, since it allows emulating human-human interaction paradigms including non-verbal communication channels, directed gaze, and even tactile interaction.

Proposition 3:

Using non-verbal signals of a robotic user interface to interrupt and alert a group of users is an intuitive way to generate subtle-but-public alerts for mobile communication devices.

I believe that embodying the Intermediary in the form of a small animated robotic animal located close to the user—sitting on her shoulder, in her chest pocket, wrapped around her neck, or placed in front of her on the desk—would allow co-located people to easily associate it with its user. This concept of a 'personal companion' that is always close by is well known in fiction literature (e.g., Pullman, 1995) [161], and can be easily applied to the telecommunication domain.

Having a physical embodiment of the Intermediary also emphasizes the intended perspective of an independent entity. In addition, anecdotal evidence suggests that disguising a cellphone as a cute animal increases its acceptance to co-located people[1].

[1] http://www.cellbaby.com

But most importantly, in the domain of personal mobile telecommunication, embodying the Intermediary in a physical entity has the following two main advantages:

- Socially appropriate communication behavior
- Focal point of attention

3.3.1. Socially Appropriate Communication Behavior

The physical embodiment allows the Intermediary to act in a socially appropriate way by bypassing socially intrusive (and annoying) alerts like ringing or vibration to interrupt and communicate with the user. Instead, it will try to catch the people's attention with subtle non-verbal behavior like opening its eyes, turning its head, wiggling its tail or ears, etc. All these cues fall into the category of subtle but public alerts. These cues are not only intuitive to humans, but they also allow for a more expressive alert scheme: the cues can vary according to who is calling, how important the issue is, or even according to how relevant the call might be for the current social setting (e.g., where the user is, who is close by, etc.). It is also possible to add more intrusive alerts to this scheme, in case the user misses the subtle alerts of an important interruption.

3.3.2. Focal point of attention

The second reason for physically embodying the Intermediary is the following: If such an Intermediary is located in the proximity of the user or even worn by the user (on ones shoulder, chest, etc.), the user can turn to it, listen and talk (or whisper) to it, making it a natural and obvious focal point of attention. This is important for the people who are engaged in face-to-face interactions with the user. Nowadays, the user's cellphone, hidden in his pocket, vibrates upon an incoming call, and the user is forced to explain explicitly that he intends to shift his attention to the phone call, if he wants to avoid confusing co-located people by just getting up and leaving the setting without explanation. Having a natural focal point of attention would avoid confusion of this kind, since it is very clear that the user just became involved in an interaction with his Intermediary.

Brooks (2003) [23] notes that embedding voice telecommunication functionality (both synchronous and asynchronous, see section 3.4.3) into a physical embodiment will have the consequence that users do not only talk *through* their communication devices (as they do today), but rather *to* them, addressing them directly. Although this might be currently perceived as a social stigma ("He is talking to his stuffed animal!"), it emphasizes the point made earlier that the Intermediary will go beyond the current paradigm of a mobile phone as a mere portal to other people.

3.3.3. Embodiment design space

Although the main reasons for embodying the Intermediary in an animatronic device are allowing the Intermediary to use human-style non-verbal cues for socially appropriate alerting and providing a natural focal point for attention for user and bystanders, I foresee that the embodiment of an Intermediary will assume a far wider range of functions for the user.

First of all, like current cellphones, it may become a *status symbol* for the owner. This implies that personal preferences for certain embodiments are influenced by the same principles that make people prefer certain consumer products to others. Assuming that the embodiments are functionally equivalent to each other—e.g., they all use non-verbal human style cues for alerting and interruption, a user's preference may be influenced by fashion trends, and—before such trends exist—by purely emotional factors.

As research in designing 'seductive products' shows, such emotional preferences can happen on three levels: reactive, behavioral, and reflective (Norman, 2004) [150]. The reactive level is concerned with appearances, the behavioral level with effectiveness of use, and the reflective level considers the rationalization and intellectualization of a product.

A preference for a cute embodiment may result from a user's emotional choice on a reactive (or visceral) level: the visual and tactile appearance of the animatronics may make all the difference.

However, some people may choose an embodiment design based on emotional preferences on a different level. If the owner chooses on a behavioral level, she might prefer a particularly efficient design, such as an embodiment that is as small and sturdy as possible, and does its job of alerting in the most efficient way. It does not matter if the embodiment is cute or cuddly—only if it does its job well.

If the owner chooses on a reflective level, an embodiment may be chosen that represents certain values of the owner towards society. E.g., an animal embodiment known for its closeness to extinction may be chosen because of its symbolic value, which may exhibit the owner's environmental awareness towards bystanders. Obviously it is also possible to express less altruistic personal values, such as leadership and 'always in control' via a design chosen based on the reflective level.

There is no right or wrong, but as Aaron Marcus (2002) [126] writes: "Cuteness can become a commodity serving relentless commercialization that, in the extreme, dehumanizes user experience, driving out variation in pursuit of megahit, lowest-common-denominator success. The cult of cute is not in itself bad, but we need to be aware of and thoughtful about how to use it in moderation. (*The Cult of Cute: The Challenge of user Experience Design*, p 33)

Unlike a mobile phone that is a piece of technology and does not project either agency or animacy, the Intermediary embodiment is designed as a 'personal companion' of the owner. It can have highly expressive zoomorphic or anthropomorphic features which may evoke much more complex reactions from bystanders. By having a preference for one embodiment over another, the owner may implicitly show that she feels this technology reflects her values, and that this artifact projects a desired image about her to others.

Although the Intermediary is not meant to be an avatar—neither for the owner nor the remote person—the owner nevertheless may use it to express certain preferred personality traits. The overall 'appeal' of an embodiment may well be more important than its cuteness—a trait that may not always be the right answer to adoption and likeability. In the fascination world of animation (Thomas et al., 1981) [198] it is well known that not only morally good characters have appeal—to the contrary, villains are often more colorful and interesting than the 'good guys' due to their unusual character traits. Appeal is the pleasing and fascinating quality that makes us enjoy looking at a character, and Thomas et al. (1981) explain: "The word is often misinterpreted to suggest cuddly bunnies and soft kittens. To us, it meant anything that a person likes to see, a quality of charm, pleasing design, simplicity, communication and magnetism." (*The Illusion of Life: Disney Animation*, p 68) [198] People enjoy watching something that is appealing to them, whether it is an expression, a character, or a movement.

Likewise, the appeal of an embodiment may easily become the single most important factor that determines personal preferences for, and differences between embodiments. For example, an embodiment reminiscent of Arnold Schwarzenegger's humanoid cult robot *Terminator*, or *Star War*'s universally dark and mysterious *Darth Vader* will allow the owner to project a distinct preference to the outside world. Both Terminator and Darth Vader are very complex characters that express determination and power, and have both crossed the border between good and bad multiple times, becoming highly controversial cult characters among the people who are interested in science fiction.

Such a projection of the owner's intentions or character traits towards the outside world is well known in fiction literature. As mentioned earlier, Philip Pullman's trilogy *His Dark Materials* (1995) [161] describes a world in which each human is born with a *daimon*, an animal companion that represents aspects of the soul of this person. Person and daimon are physically separate entities, but remain together for their whole life, and cannot be separated without both of them perishing within a short amount of time. During childhood, the child's daimon can assume a variety of different animal forms depending on the momentary mood of the young human. Once a person has reached emotionally stable adulthood, her daimon also stabilizes in a permanent form, and cannot switch between different incarnations anymore. In Philip Pullman's world, it is easy to guess if a person speaks the truth or lies because the person's daimon clearly acts out the person's inner states.

3.4. Modules of 'Residual Social Intelligence'

An Intermediary has to apply social intelligence both to its embodied interactions with humans (user interface component, previous section), as well as to its reasoning (artificial intelligence component, this section). For the former, I suggested using public but subtle non-verbal language cues performed by a small animatronic device. For the latter, I propose a set of strategies in the form of 'intelligence modules,' that allow the Intermediary to behave with social intelligence. These intelligence modules are independent ways to look at the task at hand, and are meant to be used in parallel, if available. Each module might suggest a behavior or a solution independently. The Intermediary will try to come to the best conclusion at any given point in time with whatever modules are available, using a 'fail soft' approach that assumes that a little bit of social intelligence is always better than no intelligence at all. The proposed modules are neither a complete set, nor the only possible set for human style social intelligence in communication devices: they are rather a first attempt to illustrate the design space.

The modules are based on the idea of 'residual social intelligence.' This means that the agent is not inherently intelligent as a stand-alone artificial intelligence, but harvests ‘leftover’ social intelligence from close by sources, both human and artificial. Each of the following modules relies on different resources, and represents a different perspective to the problem at hand.

3.4.1. Social Polling of Immediate Surrounding

The first intelligence module, ‘Social Polling of Immediate Surrounding,’ is based on the idea that people are socially intelligent beings. Most humans know well what is socially appropriate in a given situation, especially how to interrupt a conversation when something important comes up, and *not* to interrupt when it’s not important enough. Furthermore, humans know exactly what kind of social situation they are in and, e.g., if it is appropriate to take phone calls. Therefore, the first intelligence module that is available to the Intermediary enables it to ask the people that are involved in a face-to-face conversation with the user in a very subtle way, e.g., based on a peripheral awareness interface such as a vibrating finger ring, if an interruption from a mobile communication device would be appropriate. All involved people are given the possibility to "veto" an incoming communication in an equally subtle way, e.g., by touching their ring. This shifts the burden of deciding whether to interrupt away from the Intermediary and towards the humans who are actually involved in the face-to-face conversation.

Proposition 4:

Allowing all participants of a face-to-face conversation to ‘veto’ to an interruption by a mobile device unobtrusively and anonymously will increase the social acceptance of mobile communication.

Since the people who are alerted in a subtle way about a possible interruption do not know *whose* mobile communication device is about to interrupt, they are forced to think about the interruption with a non-egocentric perspective. Each involved person has to decide individually: Would an interruption right now be detrimental to the group’s interests? Since the vetoing process is anonymous and unobtrusive, the vetoing party cannot suffer from a social backlash. Current mobile devices are designed so that each person is responsible for interruptions by his or her own mobile device. This will not change, but I believe that adding a social polling infrastructure to mobile communication devices would foster the paradigm of a *socially distributed responsibility* for interruption by communication devices.

One of the advantages of this system is that a single veto is enough to prevent an interruption, which means that not only the owner of a mobile device prevent an interruption from his communication device, but any person involved in a face-to-face conversation with this user can do so. This allows distributing the responsibility for interruptions to the whole conversational group, and will prevent interruptions from devices that were accidentally left on by their respective users; thus, people who tend to forget to turn off their devices are prevented from getting into socially awkward situations.

The above scenario is based on an egalitarian approach. Variations may include modes where, e.g., all participants of a conversation are alerted and allowed to veto *except* the user who owns the interrupting device. Another variation may be that more than one veto is necessary to avoid an interruption, or even a majority. Yet another approach may be that different users have different weights in the vetoing process, perhaps proportional to their social status or position in a corporate hierarchy.

As a prerequisite for this module, the Intermediary has to know who is involved in a face-to-face conversation with the user. This is accomplished by Conversation Finder 'sub-agents,' a decentralized network of small body-worn wireless sensor nodes that provide the Intermediary with some information about her user's social state. A completely distributed decision-making process is used to detect conversations. The nodes have binary speech detectors and low-range radio transceivers, and communicate asynchronously with each other on a single channel. Each node sends out frequent heartbeat messages over radio, as well as specific messages when the user is talking, and receives messages from the nodes that are close by. The nodes independently come to a conclusion about who is in the user's current conversation by looking at the degree of time-alignment of the speaking parties. At any time, the Intermediary can query the user's node wirelessly for this continuously updated information.

In addition to providing the Intermediary with the identities of the people that should be notified and that can veto to an interruption, the Conversation Finder sub-agents also give the Intermediary a rough idea about the user's social setting, e.g., if she is all by herself, or part of a group, mainly listening to a speaker, or being the main speaker herself. Both the size of the conversational group as well as the ratio of listening to talking participants is available to the Intermediary.

Proposition 5:

A decentralized network of body-worn wireless sensor nodes that communicate when their wearers are speaking can be used to detect the conversational status of the user.

This rudimentary awareness of the user's social setting is necessary in order to make socially acceptable decisions about interruptions from a mobile communication device. It tries to substitute for the information people can get from just looking at a scene. For example, if two people have a conversation in one person's office, it may not be a good time for an interruption by a third person. Although the Intermediary cannot know this fact from looking at the scene (there is no vision involved), it may reach the same conclusion by evaluating the information from the Conversation Finder sub-agents. The Intermediary can detect a conversation between these two, and may know about their current location. Therefore, if one of the involved people gets a cellphone call, the Intermediary can intercept the call and relay this information to the caller, if the caller has appropriate clearance for this information. Depending on the relation between the caller and the called person, the Intermediary may even disclose the identity of the participants of the ongoing conversation.

Even within a conversation, the Conversation Finder sub-agents allow the Intermediary to interrupt an ongoing face-to-face interaction at an appropriate time. Since it is aware of who is talking at any given point in time—be it the user or any other participant of the conversation—, it can simply wait until a pause occurs in the conversation, or at least until the user herself is not talking anymore, and then try to take its turn. Such a feature may enhance the acceptability of interruptions from mobile devices during a conversation.

3.4.2. Room Memory

The second intelligence module, 'Room Memory,' is based on the idea that the physical location has a high influence on the communication behavior of people. For example, in a movie theater, during a show, people rarely take phone calls, whereas in a cafeteria, most of the people are willing to accept a phone call, no matter who calls or what the time is. The Intermediary could therefore ask the room it is in how people usually respond to calls in this specific location, at a specific time. The room could give back a summary of past events, having registered what people (or other Intermediaries) did in the past with their communication devices. This information, combined with other parameters such as how many people are present, helps the Intermediary decide what kind of interruption might be appropriate in a given location at a given time.

Proposition 6:

The information on how mobile communication devices are used (turned on, off, vibrate, etc.) in a room sized area with wireless automatic sub-agents that sum up communication events and can be queried for these results, allows an Intermediary to choose more socially appropriate communication behavior.

In addition to letting Room Memory just register the behavior of the users, it is also possible to pre-set the Room Memory with a default value, e.g., a restaurant owner may decide to disallow any phone calls in a certain section of his dining area.

Room Memory is a sub-agent with a low range radio transceiver as described in the Conversation Finder section, except that a microphone is not necessary, since Room Memory merely collects communication behaviors of users and Intermediaries close by, and re-broadcasts summaries of it to querying Intermediaries.

Room Memory could also provide the Intermediary with a different kind of information: since it identifies users via the unique heartbeat of their Conversation Finder sub-agents, it could also add up the presence information for a specific user over time. For example, if a person usually works in the office until 8pm, and then leaves the building, Room Memory could aggregate this information into a specific user profile and release it to its own Intermediary upon request. This would allow the Intermediary to predict in a crude way location changes. Although far from reliable, such information about the user's habits would be useful if a caller calls the user five minutes before he usually leaves the office. The Intermediary could disclose this information to the caller, in the form of "expected duration of an interaction," if the caller has enough clearance to know about this user habit. For example, the Intermediary could tell the caller that the user has only very little time, which would probably influence the interaction.

In order to further validate such guesses about the user's habits, the Intermediary could be allowed to access the user's calendar, and combine the information in there with what Room Memory knows to build up a user profile of location changes.

3.4.3. Intermediary capable of multiple concurrent conversations

The third module describes the ability of the Intermediary to engage autonomously in concurrent voice interactions, and mediating between them in real time in a useful and intelligent way.

If the user is not available for synchronous voice communication, the Intermediary can try to engage a caller in an interactive voice communication. The Intermediary could give the user a short summary of what the call is about, either after the call (communication activity summary when the user becomes available or returns), or—more importantly—even during an ongoing conversation, being a mediating party for the conversation. E.g., the Intermediary could tell a caller that the user is busy, and give the caller the option to leave a short voice instant message. The user can ignore this message, send back a reply, or decide to connect to the caller.

Proposition 7:

An Intermediary that is able to be involved in two concurrent conversations, one with the user, and one with the caller—at the same time, mediating between them—allows the user to reduce the time spent on the phone and increase the time spent on face-to-face interactions.

Therefore, the Intermediary does more than just pass messages between the user and the caller: being able to be involved in two (or even more) concurrent conversations, one with the user, and one with the caller(s)—at the same time, mediating between them—clearly exceeds the capabilities of a human secretary. E.g., it allows the user to deal with an incoming phone call in real time in an asynchronous way while still attending to an ongoing face-to-face conversation.

As a consequence, callers do not only have two communication modes available—talking to the called party directly, or leaving a voice mail message—but three. The third mode of communication consists of a conversation with the semi-intelligent Intermediary of the called person, which is knowledgeable about the person's current social status, and can pass short voice instant messages between the two parties.

The capability of concurrent conversations is a feature of social intelligence, because it allows the Intermediary to modify its interaction with the user depending on content of the call and the conversational status of the user. For example, it interrupts and communicates with the user when it is appropriate, be it either synchronously or asynchronously (voice instant message), being able to

Proposition 8:

An Intermediary that modifies its interaction with the user depending on the caller, the conversational status of the user, and the content of the call by, e.g., interrupting and communicating with the user when it is socially appropriate, being able to break down a synchronous phone call into asynchronous pieces, will increase the mobile communication options for a user.

break down a synchronous phone call into asynchronous pieces (chunks of voice instant messages).

The interaction between caller and Intermediary may be scripted on the highest level, since some communication events are time critical. Therefore, all conversations follow a path on large tree of pre-defined states and events, and caller, callee, as well as co-located people influence the branching conditions. However, the interaction could also include more event driven parts, and possibly branch off into less scripted interactions at certain times during the conversations. For example, when the called party listens to a voice instant message and records a reply, the Intermediary could fill this time with less scripted conversation, making the caller's wait online worthwhile. In order to do so, the conversation during these 'holes' in the scripted interaction has to be either personalized or related to the user and/or the caller (the latter being much more difficult than the former). It could include the following subjects:

- Personalized news, music, and quote of the day, suggested by the owner of the Intermediary: "Here is an article that the user found interesting." "Here is a piece of music the user is listening to quite frequently."
- Reveal to the caller more about the state of the user: First, the Intermediary might just say "Sorry, he is busy", etc. But later, during a waiting period, it could say: "He is in a conference room talking to his boss for a while already." And: "He heard your message, and is recording a reply."

In any case, there must be a good reason for the caller to stay online and wait for the user's reaction. Although such a 'chat' will be likely not very sophisticated because of low speech recognition accuracy, rigid interaction scripts could be 'softened up' to make them more flexible, and the conversation would flow more naturally, making the interaction experience for the caller more socially acceptable.

The notion of an Intermediary capable of multiple concurrent conversations could have great potential in the long run. Once users are comfortable with the basic concept of an Intermediary, the idea can be easily lifted to other areas of communication, e.g., when the two parties do not speak the same language (*interlanguage intermediary*). Since an Intermediary can downgrade a synchronous communication to an asynchronous one (voice instant messaging), the additional delay of language translation will be acceptable.

Extending the concept of an Autonomous Interactive Intermediary even further, Intermediaries could be built for interactions between human and non-human entities. This may include Intermediaries as interfaces to complex technologies such as houses, cars, and spacecrafts—but also to other species such as pets (*interspecies intermediary*). They all could have their own Intermediaries that can speak to a calling party semi- or asynchronously, being knowledgeable about the owner's 'state' and 'goals,' and translating the caller's voice communication to the other party's specific language. (For more examples, see also section 7.3.4.)

3.4.4. Issue Detection

The fourth intelligence module, 'Issue Detection,' tries to harvest information from the calling party. It accomplishes this by comparing two sources of information:

On one hand, it engages the calling party in a conversation using speech prompts and speech recognition in order to get a basic idea of what the call might be about. On the other hand, it creates and continuously updates a representation of the user's interests, the 'issues' that might be on her mind. The user's short-term interests are harvested from her ToDo list, from recently sent email, recently made web searches, and recently edited documents. Information about her long-term interests may come from analyzing the user's personal home page and other publicly available information about her on the Web. These interests and issues are represented as a simple 'bag of words.'

The Intermediary then assesses the relevance of a call to the user by comparing recognized words of its conversation with the caller with what it knows about what is currently on the mind of the user.

However, most often there will be no direct mapping between the 'bag of words' and the few correctly recognized words from the speech recognizer. Therefore, the Intermediary may try to connect the pieces of information it gets from an interactive phone call with a set of fuzzy inferences with its model of the user. In its simplest instance, the bag of words could get extended with synonyms from WordNet (Miller, 1995) [138]. In a more sophisticated approach, a query extension can be made using a semantic network like *ConceptNet* (Liu et al., 2004) [117] that is mined from *Openmind*, a large repository of commonsense knowledge (Singh, 2002) [192].

Depending on the results of such a comparison, the Intermediary might take several actions, including: alerting the user of the call immediately, telling her who is on the phone; giving the caller the option to leave a short message that is delivered immediately (voice instant message, or other modes of communication); summarizing the call itself to the user with spoken language and/or simple non-verbal behavior; suggesting an alternative way for the caller to reach the user; etc.

Proposition 9:

An Intermediary that is able to assess the relevance of a call to the user by comparing recognized words of its conversation with the caller with what it knows about the user's current interests, will allow the Intermediary to adjust its conversational script towards the caller, and make a more socially appropriate decision about when and how to interrupt the user.

The fifth intelligence module is different from the other four in that it relies heavily on recognized content of a conversation. This is difficult to accomplish when the content comes from noisy speech recognition transcripts, where recognition rate can go as low as 20%. However, it may be worth having, for the following reason:

In recent years, there has been a backlash against context aware systems and systems using agent and artificial intelligence approaches. Some researchers have given up on trying to create 'truly' intelligent systems, because they came to the conclusion that it is too hard, that they cannot be built (Erickson, 2001) [56]. Instead, these researchers retract to other strategies like giving people more low-level context information so that humans can develop context awareness themselves.

This situation seems reminiscent of the early days of artificial intelligence when researchers found out that building computers with generic human knowledge is too hard, and retracted to sub-problems of intelligence and to narrowly focused practical A.I. projects, like chess playing. Decades later, it starts to become clear that it was not a good idea to fall back on easier problems, because the bigger ones would not go away. E.g., even today, no genuinely intelligent 'thinking machine' can be built from the elements of currently available A.I. technologies. As a consequence, researchers like Minsky try to tackle the original problem as a whole again, e.g., via the Commonsense Reasoning approach, instead of building specialized systems that are very brittle (Minsky et al., 2002) [141].

The same may happen to context aware systems: falling back on trying to solve simpler problems will not solve the big problem. Therefore, it may be worthwhile trying to develop an intelligence that does not rely on human interpretation of lower level sensor data. Such as system may solicit human input ('human augmented A.I.'), but *should come to a useful idea about the world without people having to interpret the available sensor data*.

Summary

This section described an Intermediary that controls a user's mobile communication devices. It has not only non-verbal expressive capabilities to attract attention and communicate with the user and co-located people in a public but subtle and socially appropriate way, but can also rely on intelligence 'modules' that allow it to harvest 'residual social intelligence' from its surroundings, such as:

- unobtrusively polling the co-located people for advice about how to behave in a socially appropriate way
- querying the room's memory for how mobile communication has been handled at this location in the past by humans and their Intermediaries
- concurrently interacting with the user and the calling party in order to mediate between them in real time
- understanding the relevance and importance of a mediated call by trying to make semantic connections from recognized words of the call to a set of issues that might be on the user's mind

3.5. User experience of ideal Intermediary

The following four stories illustrate what the user experience of an ideal Autonomous Interactive Intermediary may be.

The current implementation of an Intermediary includes many, but not all of the features that are described in these stories. The not yet implemented features are marked with a star *.

3.5.1. Before a meeting

John is about to have a work related meeting with his partners. They are sitting in a conference room, waiting for another participant. They are chatting about vacations and their newest gadgets. Hillary, John's wife, calls him on his cellphone. The Intermediary that controls the cellphone knows from its sensor network nodes that John is in a conversation with five people. The caller is important—it knows that it's his wife from caller ID—, so it wants to alert John immediately, but still asks the conference room's Room Memory how often people take phone calls in here. The room tells the Intermediary that a relatively high percentage of the people have their phones usually turned off, especially during the day when there are frequent meetings. Therefore, the Intermediary hesitates to ring the phone, and decides to ask the people around John for advice. It queries the sensor network nodes about who John is talking to right now, and alerts the people with a slight vibration of their actuated finger rings of an incoming call. The involved people's rings vibrate, alerting each of them in a subtle way of a possible interruption, and at the same time allowing them to disable an incoming call. But nobody has any objections right now since the meeting has not started yet, and so everybody ignores the pre-alert. After a few seconds, the Intermediary rings John's phone (there is no need for more subtle alerts at that point since nobody seems to care about the interruption anyway), and he picks it up and talks to Hillary.

3.5.2. During a meeting

The meeting has started. It is important. John's cellphone, in the shape of a cute little stuffed animal, is sitting right in front of him on the desk (as are all the other participant's Intermediaries). It is asleep, eyes closed, just slightly and silently breathing. Hillary calls again. The Intermediary knows that John is in a conversation, and that he is in a conference room. Again, it vibrates all the available participants' finger rings. This time, somebody immediately touches his ring to disable the call—still not knowing whose communication device is receiving a call. Therefore, the Intermediary takes the call and tells Hillary that it is aware of her importance, but it thinks that the social setting does not allow John to take any calls. It asks her to explain briefly what this is about. She says that the landlord has called about the new heating. The Intermediary recognizes the words “heating” and “landlord,” and makes a link to an entry of John's ToDo list*, which says “Ask Mr. Gilliam to look at the heating.” “Heating” in itself might not be enough, but it also knows that Mr. Gilliam is in John's address book as “landlord,” and has grouped the words “Gilliam” and “landlord” together*. At that point, the Intermediary regards the call as important enough to try to get John's attention, and tells Hillary to stay on the line. The cute little animal on John's desk wakes up, opens its big eyes, and looks around. John sees it, but ignores it—he is too busy, even for a phone call that might be important. After a few seconds, the creature gives up, shakes its head, and goes back to sleep. The Intermediary tells Hillary that John did not respond, and asks her if she wants to leave a voice mail. She does so. After she is done, the stuffed animal gives John a little hint that the calling party left a message via a little twitch of its head.

3.5.3. At home

John is at home, eating lunch with his family. His cute creature sits in his chest pocket*, sleeping, as usual. John's friend is calling, but the Intermediary does not recognize the caller ID, since he's calling from his parents' phone. The Intermediary knows where John is right now, and that it is lunchtime, and that he is talking to his family*—not a good time to take phone calls. Since it doesn't know who is calling, and that John is sitting at the lunch table, it doesn't bother polling the rest of the family via the actuated finger rings—the call's not important enough at that point. It takes the phone call instead of John and asks who this is and what this is about. John's friend says this is Mike, and that he can't attend the fishing trip they have planned. But the Intermediary does not recognize any significant words, so it tells the caller that it is very sorry but it couldn't bother John at that point, mentioning also that the phone connection is very noisy and it has a hard time understanding him*. The Intermediary then suggests that the caller leave an instant voice message and stay on the line for a short time, if he wants to. John's friend agrees and says "It's me, and I can't come to the trip tomorrow. Please pick up the phone!" The Intermediary in John's pocket wakes up, opens its eyes, and opens its mouth, like it is about to say something*. The family recognizes the non-verbal signals of the animal, and they start to talk about something that does not include John. So John looks down on his Intermediary in his pocket and says in a very low voice: "Ok, what is it?"* The animal says, with the same very low voice: "Short voice message from a caller on hold. You want to hear?" John whispers "Sure"*, and the Intermediary plays the message. John thinks the issue is important, but doesn't want to talk to his friend right now, so he grabs the ear of the stuffed animal (which starts the recording process), and says: "I will call you back in an hour." As soon as he lets go of the ear, the Intermediary tells the caller on hold: "John can't take the call, but he left a message" and plays back the message.

3.5.4. Concurrent conversations

John sits in his office. He expects two visitors. His Intermediary is sitting on his shoulder*, sleeping, as usual. His visitors arrive and take seats. They get into a lively discussion about the future direction of John's work. Suddenly, the Intermediary wakes up, and wiggles its tail*. A few seconds ago John's father, Curt, called. The Intermediary intercepted the phone call and told him that his son is talking to two people. This is the complete interaction (vertical axis is time):

Curt (calling party)	*Intermediary*			*John (called party)*
Dials John's number				
Ringing Ringing Ringing	Gets caller ID Detects John being in a conversation, talking to two people in his office			
	Hi Curt, John is talking to two people in his office. Would you like to leave a message?		Opens eyes, wiggles tail.	
				(Turns around, whispers) Yes?
			(Whispers) Your father on the phone!	
				Hang on a second.
Yes. I am coming for dinner tonight, and was wondering if I should bring some wine. (Pause)				(Continues talking to visitors…) (Keeps talking to visitors…)
	Ok, I will tell him as soon as possible. Would you like to wait?			
Yes. How long will it take?				(Turns to Intermediary and touches its nose)
	Sorry, I didn't understand the second part of what you just said. Could you repeat that please for me?			
			(Whispers) He left a message: "I am coming for dinner tonight, and was wondering if I should bring some wine."	
That's ok. I can wait a little. Never mind.				
	I just told him—waiting for his answer...			(Touches ear) Tell him: That would be great! (Lets go of ear)
	John says: "That would be great"		(Whispers) I will tell him. Anything else?	
Ok. Thanks, bye!				Nope!
	Bye!		(Falls asleep again)	

An Intermediary being able to lead concurrent conversation like in the above example allows the user to interact briefly and politely with people with whom he would otherwise chat with for extended periods, if they both had time for it. Cutting these people short in a synchronous conversation, however, could be interpreted as rude and unacceptable. The Intermediary provides an elegant way out of this social dilemma.

4. Implementation

My implementation of an Autonomous Interactive Intermediary consists of computer hardware (PC level, microcontroller level, other electronics), software (Perl, C++, VB, Python, C, assembly code), a variety of radio transceivers (433Mhz, 2.4GHz), and animatronic parts (servos, sensors).

4.1. System Overview

Ideally, an Intermediary is a *completely autonomous, self-contained entity.* It is meant to be a permanent companion of the user: wherever she goes, the Intermediary is with her. An Intermediary can be carried or worn, but in order not to bother the user, it may not be larger than the size of a cellphone.

Although an Intermediary incorporates cellphone functionality, it goes far beyond what a cellphone is capable of today. Although there are cellphone platforms that do speech recognition, the Intermediary's dual conversational agent is too computationally demanding to run directly on a phone. Furthermore, cellphones today incorporate neither animatronic elements, nor connect to sensor networks. My Intermediary does both, and more.

For these reasons, it was decided to run the Intermediary's computationally intensive processes on a desktop computer. The actual agent software runs on this computer and communicates with the Intermediary's embodiment via a wireless audio and data link. This approach is commonly referred to as "remote brain robotics," and has proven to be very successful in order to test paradigms and implement functionality that cannot be implemented locally on a platform with restricted resources. However, the ultimate goal is to run all agent processes on the user's phone and control the embodiment via short-range wireless link, or alternatively to integrate phone and embodiment into one device altogether.

But even when cellphone and animatronics can be integrated and miniaturized into one tiny device, the Intermediary still relies on a sensor network that cannot be part of the cellphone itself. Ultimately, each person may wear one or several tiny sensor nodes, either in the shape of jewelry (including wrist bracelets, belt buckles, rings, etc.), or sewn directly into the clothes. These nodes will form an adhoc and completely decentralized sensor network that will serve as a shared resource for all Intermediaries in proximity.

My Intermediary consists of the following main subsystems:

- **Remote computer:** located within range of audio and data transceiver; runs all high-level control processes; has a landline phone interface; runs speech recognition server; access to wireless data transceivers (for animatronics and sensor network)

- **Animatronics:** to be carried or worn by user; sensors and actuators controlled locally by microprocessors; wireless duplex audio and data link to PC for audio functionality (cellphone) and to relay actuator and sensor data
- **Conversation Finder nodes:** to be worn close to the neck; overall size less than 40mm
- **Finger Ring nodes:** to be worn on finger

Figure 2 shows an overview of these subsystems.

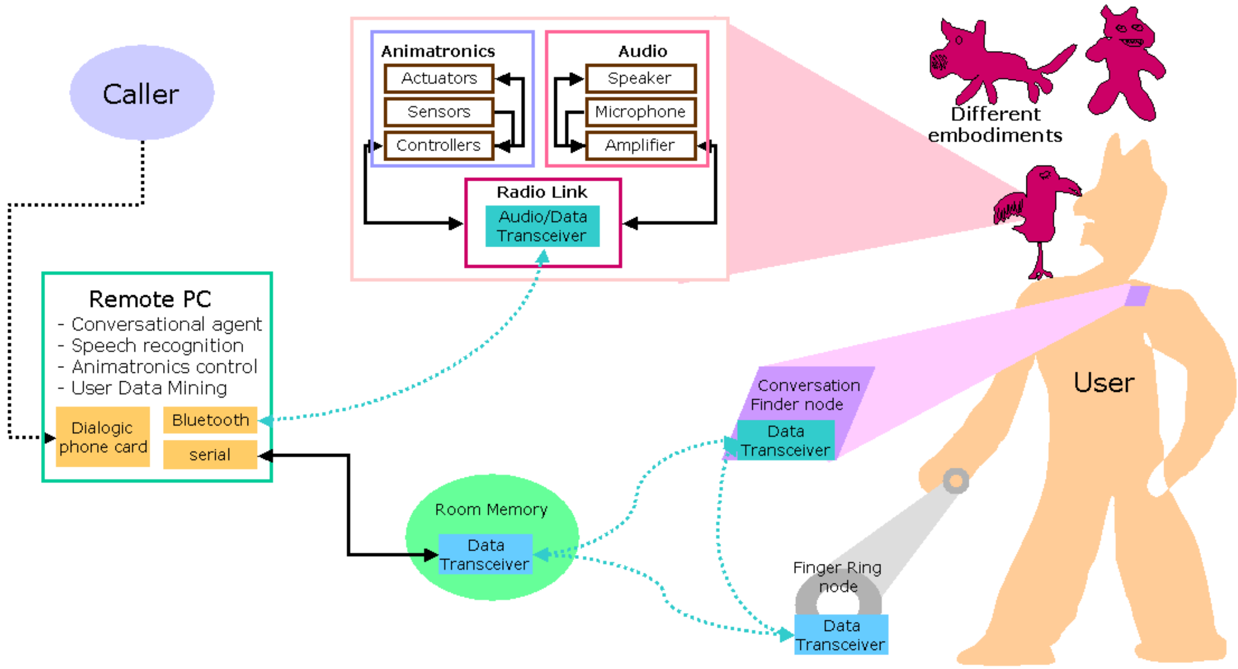


Figure 2: Architecture overview of the Autonomous Interactive Intermediary implementation, showing its subsystems

What follows are short descriptions of the Intermediary's subsystems.

4.1.1. System components

Remote computer
There are two computers that run the system's main processes:

Computer 1 (Windows):
Hardware:

- Phone interface (Dialogic card): landline call control (4 lines)
- Bluetooth transceiver: audio/data communication with animatronics
- Data transceiver: communication with sensor network

Servers:

- Conversational agent: interacts simultaneously with caller (on the phone) and user (via Bluetooth audio in the animatronics).

- Speech recognition server (Microsoft Speech)
- Animatronics control server
- Sensor network bridge server

Computer 2 (Linux, just servers):

- Data mining processes: PERL scripts collecting user information (from IMAP server, from Google API, etc.)
- Commonsense tools (ConceptNet) for fuzzy query extensions

Animatronics
There are several different instantiations of enhanced stuffed animals. Overall size of the creatures is between 11cm and 30cm.

Actuated degrees of freedom include eyes opening/closing (bunny, squirrel), looking up (bending neck back) or uncurling (from curled position to straight back), turning head, and wing movements (parrot).

Animatronics control server
This software, running on a remote PC, receives high-level messages from the conversational agent and sends servo signals to the animatronic device via Bluetooth wireless serial data link.

An earlier prototype (bunny) included an R/C handset (Futaba, 6 channels), interfaced with a modified iRX[2] "glue" board to the serial port of the PC. On the receiving side, micro R/C gear was used, such as Cirrus 4.4 micro servos and Cirrus micro receiver.

Bluetooth transceiver
The audio and data transceiver system is a Bluetooth class 1 dongle, extended with an external antenna, which resulted in an indoor range of about 40 meters, covering completely a floor of the MIT Media Lab.

The Bluetooth transceiver provides a wireless duplex audio and data connection between the animatronics and the PC that controls the speech prompt playback, speech recognition, as well as phone control. It basically extends the PC's audio in and out to the (mobile) animatronic device. On the animatronics, a Bluetooth transceiver board is connected to a small audio amplifier, speaker, and microphone.

The Bluetooth transceiver also provides a duplex data channel. Via this serial channel, the animatronics receives high-level servo control signals from the animatronics server, and simultaneously sends back the animatronics' sensor data.

Animatronics controllers
In the animatronics, there are two microcontrollers (PIC 16F84A): the first one reads the switches in the animal's extremities, and sends back the status of each switch via Bluetooth channel. The second controller receives the serial servo data from the Animatronics control server, and generates the pulse width modulated (PWM) signals for the servos.

There are three switches in the extremities of the animatronics. They are generally used as Yes, No, and Connect/Disconnect buttons, but

[2] http://web.media.mit.edu/~ayb/irx/irx2/

their functionality varies slightly depending on the status of the animatronics. In earlier embodiments (bunny), there was an additional switch in the creature's ear, which was used as a push-to-talk button.

Conversation Finder nodes
Each node consists of two double-sided PCB boards with two PIC 16LF877 controllers, microphone capsule, Radiometrix Bim2 transceiver (433MHz), microphone pre-amplifier, and 140mAh lithium polymer battery. The overall size of a node is 40x35x20mm.

Each user owns his or her Conversation Finder node, worn close to the neck. It functions as binary speech detector and communicates asynchronously with other nodes on a single radio channel. Each node sends out frequent heartbeat messages over RF, as well as specific messages when the user is talking, and receives messages from the nodes in proximity (approx. 10 meters). Each node independently comes to a decision about who is in the user's current conversation by looking at alignment and non-alignment of the speaking parties. At any time, the Intermediary can query the user's node wirelessly for this continuously updated list of people, as well as for other information concerning the user's conversational status.

Finger Ring nodes
The actuated ring consists of a tiny vibration motor (pager motor with an eccentric weight), a 20mAh lithium polymer battery, a micro switch, a Radiometrix Bim2 transceiver (433MHz), and a 16F877 microcontroller.

The Finger Ring's transceiver receives messages from its user's Conversation Finder node when it has to vibrate, upon which it vibrates slightly. If the user touches the micro switch located under the ring, the transceiver broadcasts an anonymous veto message to the Intermediary. For user testing, wired versions of the Finger Ring were developed.

Room Memory nodes
Room Memory nodes are implemented as virtual nodes in software, and use the sensor network base station with Radiometrix Bim2 transceivers.

4.1.2. System communications

In this section, I will describe how the main system components communicate with each other. I will distinguish between two system states:

- Upon system startup
- Upon incoming call

Upon system startup
In order to start up the Intermediary, several connections have to be established in a certain sequence (see Figure 3):

- The sensor controller on the animatronics goes through a sequence of serial commands to set the Bluetooth board into duplex audio and duplex data mode. The Bluetooth board attempts a Master-Slave connection to the Bluetooth dongle on the remote PC.

- After this sequence, she sensor controller starts to read the positions of all switches and generates serial signals that it sends to the Bluetooth board.
- The Bluetooth board sends back this data to the animatronics control server via the wireless link.
- As soon as the animatronics server reads sensor signals from the serial port, it sends a socket message to the conversational agent software that a connection to the animatronics has been established.
- The conversational agent receives this message, and sends back its first high-level command "System Stand by."
- The animatronics server looks up the primitive behaviors associated with "System Stand by", and starts generating the basic serial signals for the servos.
- The servo signals from the animatronics server are sent over the Bluetooth serial data link to the Bluetooth board in the animatronics.
- The servo controller board reads these serial signals and generates a continuous PWM signal for each servo.

At this point, the system is up and running.

The communication protocols between the subsystems will be described in greater detail in later sections.

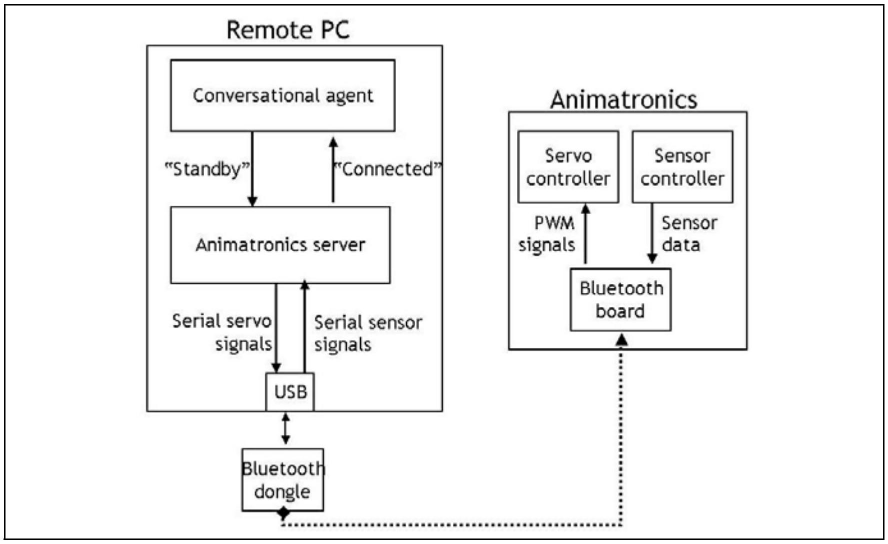
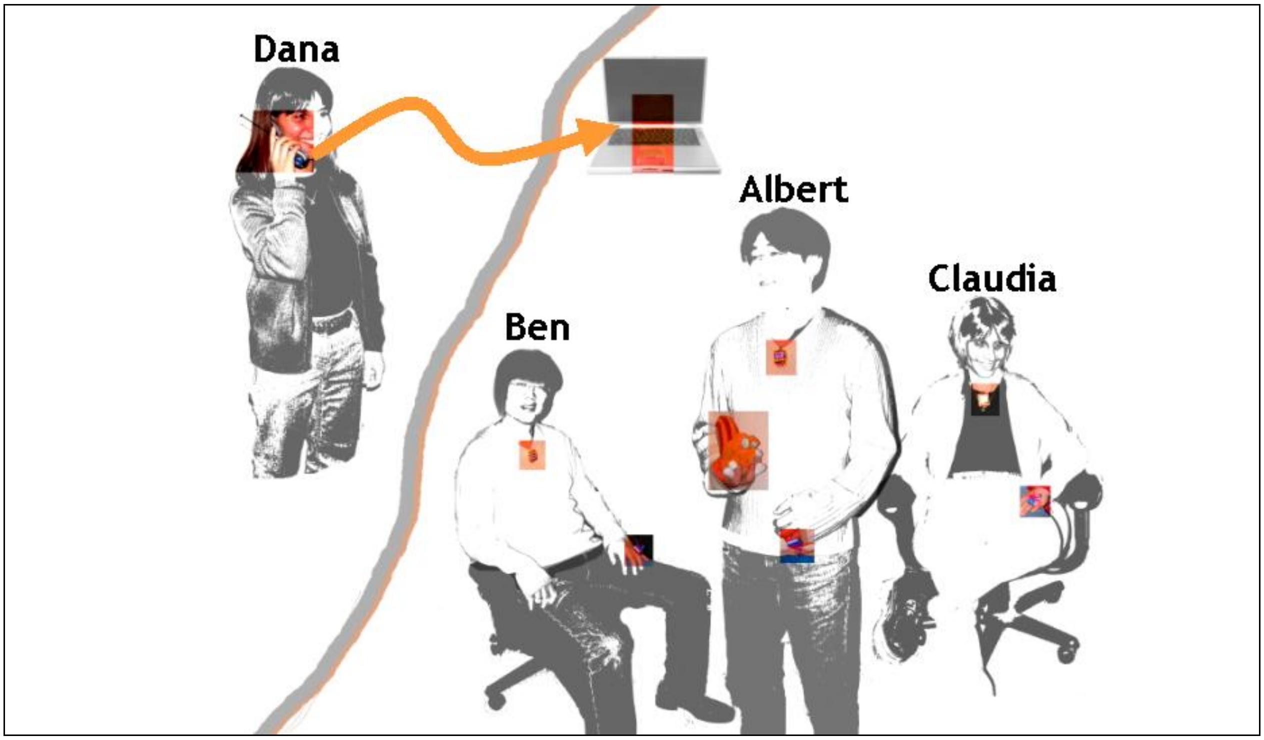


Figure 3: System communication at startup time

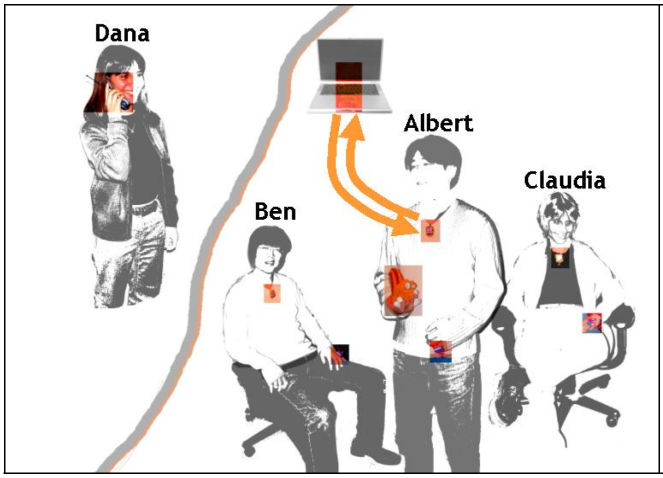
Upon incoming call

When the Intermediary receives a phone call, it first contacts the sensor network to establish the conversational setting of the user via the Conversation Finder nodes. In a second step, if necessary, it polls all conversational participants for their input via the Finger Ring nodes.

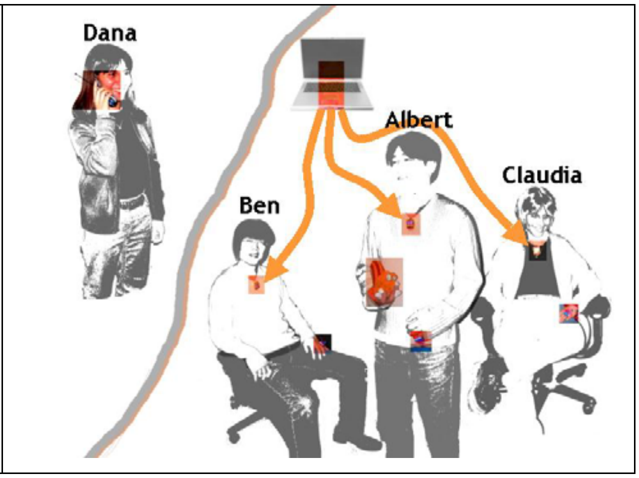
The following figures illustrate the communication between conversational agent and the sensor nodes.



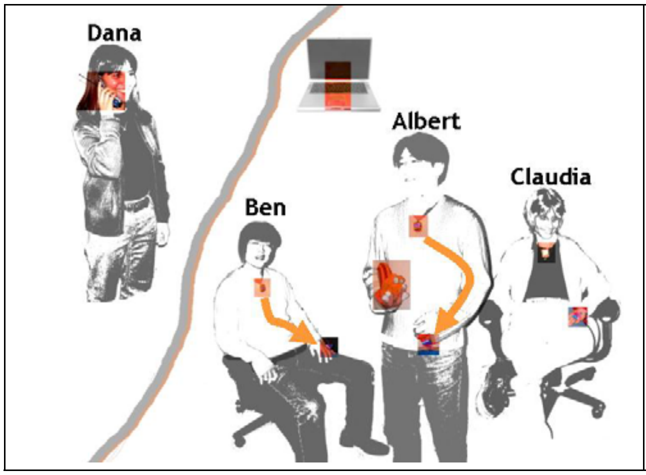
The setting is as follows: **Albert** is in a face-to-face conversation with **Ben**. They are in the same room as **Claudia**, but she is not part of their conversation. All participants wear Conversation Finder nodes as well as Finger Ring nodes. Albert is holding his Intermediary, a squirrel, in his hand. **Dana,** who is at a remote location, is calling Albert. The conversational agent, running on a remote computer, registers the incoming call for Albert.



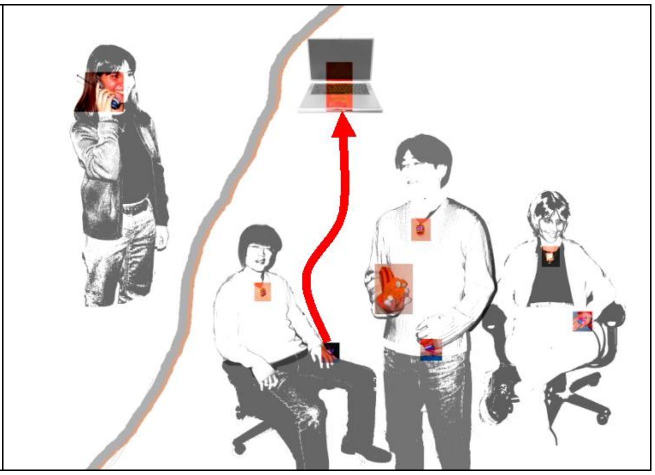
The agent first determines Albert's conversational status. It sends a socket message to the sensor network bridge server. The server sends an RF message to Albert's conversation finder node, asking how many people are in his conversation, and how much he has been talking recently. The node sends back the requested information.



In a second step, the agent polls the conversational partners of Albert. It broadcasts a message to all Conversation Finder nodes in range: If they think they are in a conversation with Albert, please notify their users of the upcoming call! All three Conversation Finder nodes (Albert, Ben, Claudia) receive the message.



However, only the nodes of Ben and Albert think they are in a conversation with Albert—Claudia's node does not think so, since it registered her talking at the same time as Albert for several seconds. Ben and Albert's nodes send messages to their respective finger rings to vibrate. These two finger rings vibrate shortly.



Ben notices the pre-alert, and thinks it is inappropriate to get an interruption right now, so he touches his ring slightly. The ring broadcasts an anonymous veto message, saying that it vetoes to the interruption by Albert's agent. Albert's conversational agent receives the veto, and takes it into account when deciding if it wants to interrupt Albert.

4.2. Conversational Agent

The previous section briefly explained the interaction of the conversational agent with the sensor nodes. This section will describe in detail the workings of the conversational agent.

From the perspective of the human user, the Intermediary consists of two types of 'agency':

- Embodied agent: for the owner and co-located people
- Conversational agent: for the owner and the calling party

The former will be discussed in section 4.4, and this section will address the latter.

For a caller, the conversational agent may appear first as an ordinary answering machine or voice mail system: it picks up the call instead of the user. Indeed, the Intermediary is intended to eventually make answering machines and voicemail obsolete and is perfectly able to 'emulate' such systems. However, the Intermediary transcends the capabilities of an answering machine in several ways. For example, it has the capability to mediate between caller and user in real time, being able to converse with both parties at the same time. It is also superior to a voicemail system because it takes into account the current conversational status of the user.

4.2.1. Call tree

The conversational agent, implemented as a finite state machine (Figure 4), follows a decision tree with branches that depend on external data and sensors, as well as caller and user choices, which are detected via speech recognition and tactile feedback. The following are the main factors influencing state changes:

- Distinction between known and unknown callers via caller ID and a list of known callers
- Caller and user choices: using speech recognition, both caller and user may choose between different modalities including voice mail and voice instant messages, or may choose to ignore the partner
- Knowing if the recipient of the call is engaged in a conversation
- Getting input from others in the co-located conversation
- Knowing how other people in this location have responded to incoming calls

When a call comes in, the Intermediary first polls the user's conversational size and determines how often she spoke recently (section 4.6). If she is in a conversation with somebody, or she talked for more than 25% during the last 15 minutes, the Intermediary assumes that she is busy. If she is not busy, however, the conversational agent plays a ringing tone and connects the caller directly to the user, which results in a full-duplex audio connection between caller and user.

If the user is busy (as defined above), the Intermediary polls all participants of the co-located conversation by asking their conversation finder nodes to vibrate their finger ring nodes. All participants then have a 10-second window to veto anonymously to the call (section 4.7).

During this window, the Intermediary keeps collecting information, such as caller ID, and compares the ID with a list of known people. Then it greets the caller, and asks her if she wants to leave a voicemail message, or needs an immediate response. If the caller chooses voicemail, the system records the message and terminates the call.

If the Intermediary recognizes the caller from caller ID and the caller needs an immediate response, the Intermediary lets her record the message, alerts the user, plays back the message, waits for a reply, and plays back the reply to the user. However, if the caller is not known, the conversational agent asks her first for more details about the call and her identity. The caller's answers are recorded and fed into the speech recognition engine, which is loaded with a specific vocabulary that tries to detect certain keywords that might be of interest to the user (section 4.9).

If the caller mentions a certain amount of interesting keywords, the conversational agent moves on and lets her record a voice instant message, and follows the path described above.

At any point in the conversation, the owner has the possibility to influence the caller's mode of communication by interacting with her animatronic device. If the user presses the front paw, the caller gets connected directly to the user, regardless of the caller's previous choices. If the user pressed the animatronics' back leg, the caller gets sent to voicemail immediately—regardless of the caller's choices. In each of these cases, a short prompt is played to explain the situation to the caller.

Similarly, if one of the co-located people vetoes to the call (within a 10-second window), the caller gets sent directly to voicemail.

The idea is that there is a clear hierarchy among all involved parties in terms of communication mode changes. The hierarchy is as follows:

1. Owner of the Intermediary
2. Co-located people
3. Caller

The conversational agent first checks the highest priority source, the owner of the Intermediary. She can influence the call at any time by interacting with the animatronics. Her choices are equivalent to "Connect the caller through!" (picks up the phone), and "Do not bother me now!" (unplugs the phone).

Below the user in the hierarchy are the co-located people. They can influence the call tree by vetoing. If the user does not express any preferences, the Intermediary checks if it has received valid vetoes. If it did, the caller is sent to voicemail directly.

And finally, the conversational agent takes into account the preferences of the caller by evaluation her spoken language choices via speech recognition. Both the owner of the Intermediary, as well as vetoes from co-located people can override her choices, though.

Although the caller has the lowest priority of all parties and her choices can be 'overruled' by either co-located people or Intermediary owner, there is a safeguard built into the system for emergencies that allows the caller to make sure that her call still gets through. The conversational agent supports 'barge-in,' meaning, the caller can interrupt the agent's prompts at any time. If the caller does so, the currently playing prompt is halted and the conversational agent records the callers words and sends them off to the speech recognizer, looking for special 'emergency' keywords such as 'hospital,' 'accident,' and 'death.' The idea is that there has to be a possibility for the caller to override the hierarchical command structure in cases of emergency.

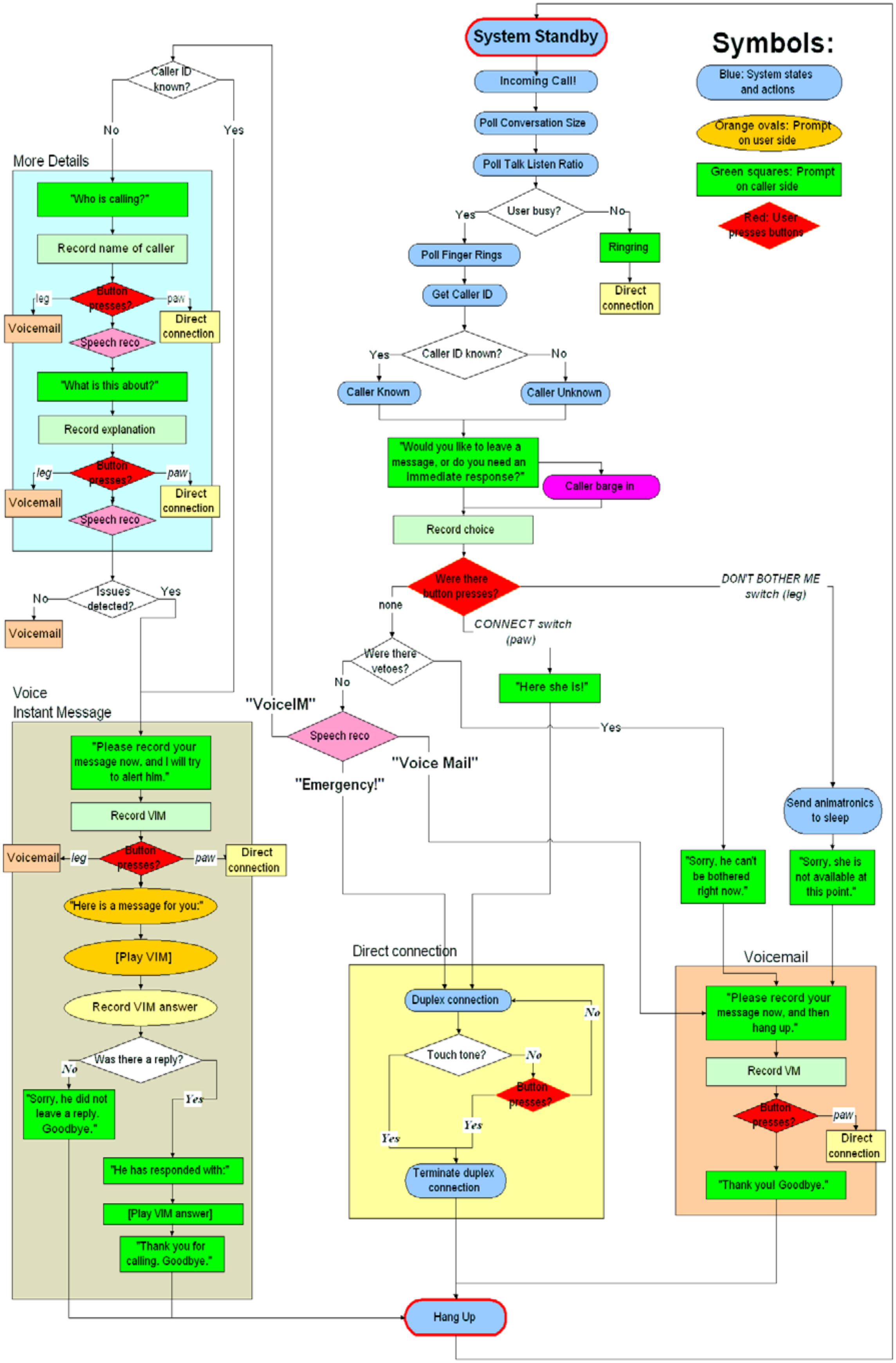


Figure 4: Call tree of dual conversational agent

4.2.2. Hardware

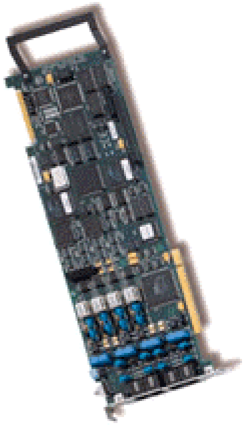


Figure 5: Dialogic phone card

The conversational agent runs on a Windows® PC, an IBM® IntelliStation M Pro 6850. This machine has a dual 1.7GHz processor, 512MB RAM, and runs Windows XP. This machine also runs most other software related to the Intermediary.

The computer hosts an internal phone card that allows the software to receive and dial phone calls. The phone card is an Intel® Dialogic® D/41JCT-LS full length PCI card (33cm long). This four-port, analog communications board is used for developing global, enterprise applications such as unified messaging, IVR, and contact centers. The D/41JCT-LS supports voice, fax, and software-based speech recognition processing in a single PCI slot, providing four analog telephone interface circuits for direct connection to analog loop start lines.

In its current implementation, the Dialogic card utilizes only a single landline, but is built to serve four. Therefore, it may be possible to run four Intermediaries on this machine, but it has not been tested.

4.2.3. Software

The conversational agent is written in C++, and its software architecture is as follows (Figure 6):

On the top level, the Main Demo code instantiates six main objects:

- *DialManager*: manages the Dialogic phone card and its low-level hardware features such as line state detection, touch-tone detection, caller ID detection, etc.
- *DialAudio*: handles audio playback and recording of the phone card; enables full-duplex conversations, pause detection, barge-in, etc.
- *SpReco_Client*: deals with the speech recognition server
- *BT_Client*: handles audio to and from the animatronics (via Bluetooth)
- *Animatronics_Client*: interacts with the animatronics server
- *Cfinder_Client*: interacts with the sensor network hub, which allows communication between conversational agent and Conversation Finder and Finger Ring sensor nodes

Some of these modules are rather complex. For example, the code that allows for a duplex audio connection between caller (from the Dialogic card) and animatronics (via Bluetooth audio device) employs a multiple buffering strategy to make sure the audio streams pass in both directions with minimal delay. In trials it was decided that a delay of 200ms is acceptable without tying down the computer’s processor too much, but still making sure that the delay does not disrupt the conversational partners.

The main modules rely on sub-modules, such as *SocketInterface.cpp,* which enables the multiple socket connections between the clients and servers, and *WaveAudio.cpp* that deals with all low-level audio functions, including a more convenient pause detection algorithm than the Dialogic's native one.

Since the agent's processes are multi-threaded and difficult to follow, the software creates an extensive log file for later analysis, which includes saving all audio messages that have passed through the system, speech recognition results, etc.

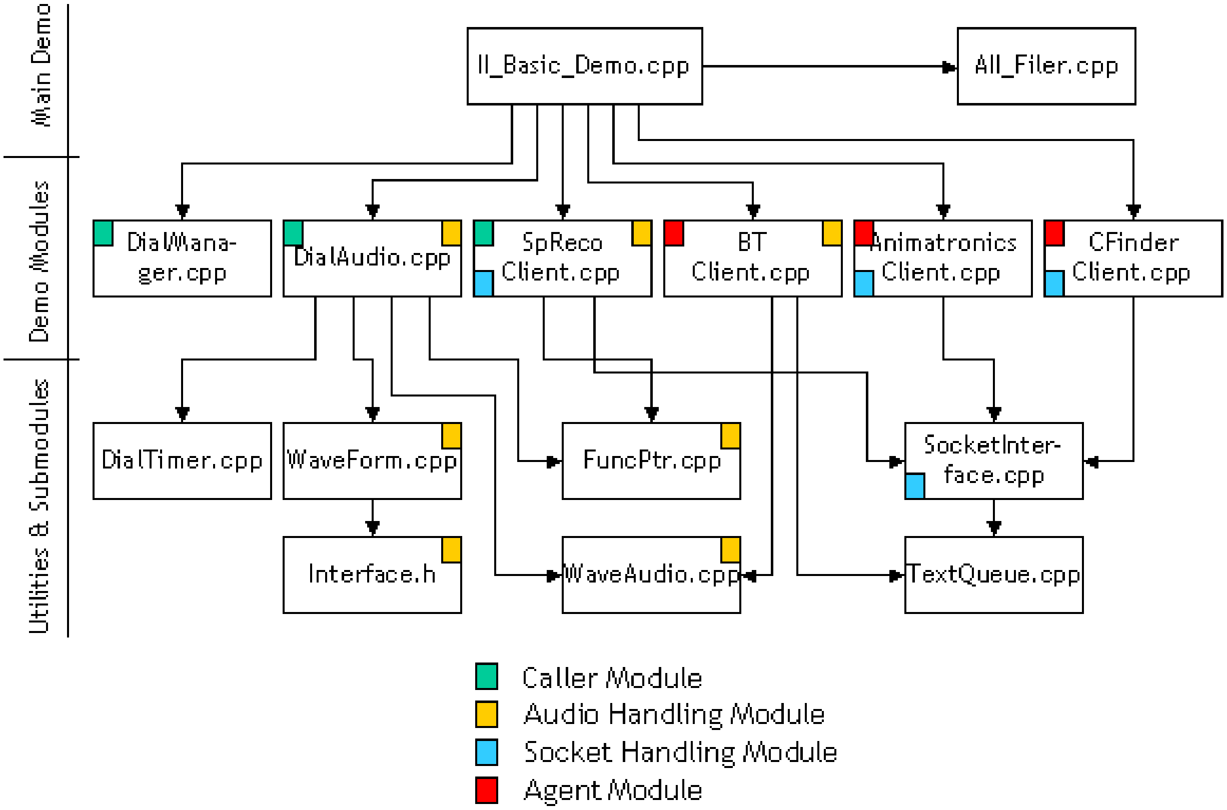


Figure 6: Conversational agent software architecture

Speech recognition
The conversational agent relies on a speech recognition server based on Microsoft Speech, sending audio buffers and getting back the recognition results. It can dynamically change the recognizer's vocabulary, which is specified as an XML file. Both the audio that was sent as well as the speech recognition output is stored for each session.

4.3. Developing the Intermediary embodiments

An important element of this thesis work is embodying the user interface for a call handling agent in an animatronic device. The embodied agent’s primary function is to interact socially, with both the user and other co-located people. Humans are experts in social interaction. We find interaction enjoyable, and feel empowered and competent when a human-machine interface is based on the same social interaction paradigms as we use (Reeves et al. 1996) [164].

4.3.1. Non-verbal cues for interruption

How do people interact with and interrupt each other? What kind of non-verbal cues are used?

Non-verbal cues are communication signals without the use of verbal codes (words). Such cues can be both intentional and unintentional, and most speakers and listeners are not conscious of these signals. The cues include (but are not limited to): touch, glance, eye contact (gaze), volume, vocal nuance, proximity, gestures, facial expression, pause (silence), intonation, posture, smell.

The problem is well studied for dyadic conversations with speakers and listeners taking turns. For example, Duncan et al. (1974) [48] show that turn-taking behavior is a complex multi-step process involving a strict pattern, which—if not followed properly—will result in simultaneous turn taking and confusion. There is a multitude of signals that are used to regulate this behavior. Of particular interest in this context are eye contact and gestures, e.g., a listener raising hand into gesture space as a nonverbal wanting-turn cue (e.g., McFarlane, 1997; Riley, 1976) [131][169].

However, an Intermediary’s task to interrupt is different from signaling turn taking in an ongoing conversation. It is rather comparable to an outside person trying to interrupt an ongoing face-to-face conversation. Experts for these kinds of interruptions are administrative assistants who are professional ‘interruption mediators.’ They make decisions every day about whether to allow interruptions to the person they support. Dabbish et al. (2003) [33] have conducted a series of interviews with administrative assistants and suggest a production-rule model of the decision process they use when deciding whether to deliver interruptions to the person they support.

Ideally, the Intermediary embodiment would learn the ‘mechanics’ of such behavior by imitating interactions between humans, perhaps starting with facial mimicry (Breazeal et al., 2005) [20]. Such a capability may well be a significant stepping-stone to developing appropriate social behavior, to predicting other’s actions, and ultimately to understanding people as social beings. However, the focus of this thesis is not on letting the embodiment develop such behavior autonomously, but to merely use human-style cues in order to alleviate the interruption problem.

In order for an agent to be understandable by humans, it must have a naturalistic embodiment and interact with its environment like living creatures do (Zlatev, 1999) [209] by sending out readable social cues that convey its internal state. It is not implied that the Intermediary’s software mimics mental cognitive processes. However, it is designed to express itself with human-style non-verbal cues such as gaze and gestures to generate certain effects and experiences with the user. The underlying idea is that human-style social cues can improve the affordances and usability of an agent system.

One of the key elements to this work is giving a conversational agent physical presence, through interactive critters of different shapes and sizes, remotely controlled by a computer. These creatures interact with a combination of pet-like and human-like behaviors, such as waking up, waving for attention, or eye contact. These non-verbal cues are intuitive, and therefore may be ideal for unobtrusive interruptions from mobile communication devices. Physical activity of the embodied agent can alert the local others to the communication attempt, allowing the various parties to more gracefully negotiate boundaries between co-located and remote conversations, and forming “subtle but public” cues as described in Hansson et al. (2001) [77]. Furthermore, these cues allow for more expressive alerting schemes by embedding additional contextual information into the alert. For example, the agent may try to get the user’s attention with varying degrees of excitement, depending on the importance or timeliness of the interruption.

The animatronics are also ‘socially evocative’ as they rely on our tendency to anthropomorphize and capitalize on feelings evoked when we nurture, care, or are involved with our “creation” (Fong et al., 2002) [57]. The embodiment serves as a social interface by employing human-like cues and communication metaphors. Its behavior is modeled at the *interface level*, so the current agent is not implemented with social cognition capabilities. Yet, it is ‘socially embedded’ since the agent is partially aware of human interaction paradigms. For example, with its capability to detect speech activity and conversational groupings in real-time (Marti et al., 2005) [127], the agent may choose to interrupt the user only when there is no speech activity.

My current embodiments are zoomorphic, but employ anthropomorphic behaviors (gaze, gestures). Although this combination partially violates the ‘life-likeness’ of the creatures, it also allows to avoid the ‘uncanny valley,’ an effect where a near-perfect portrayal of a living thing becomes highly disturbing because of slight behavioral and appearance imperfections.

Embodying an agent grounds it in our own reality. Embodiment is a structural coupling between system and agent, which creates a potential for ‘mutual perturbation’ (Dautenhahn et al., 2002) [39]. The more the system can interact with its environment, the more it is embodied.

In the current system, embodiment is realized on two levels. First, the degrees of freedom of our animatronics allow the system to ‘perturb’ its environment via physical movements. Second, the dual conversational capability that enables the system to engage in spoken interactions with both user and caller, embodies the agent in the conversational domain, which is equally human accessible. On both levels, the agent can manifest its internal state towards its environment (the caller, the user, and co-located people), and get input from its environment (spoken language, tactile) via its sensors and actuators. For example, the embodiment changes its movements when there is an incoming call, further differentiating between known and unknown callers using non-verbal signals to ‘act out’ what is going on in the phone domain.

The current embodiments are all based on animals (bunny, squirrel, and parrot), but their respective morphologies are diverse enough so that their appearances create different expectations (and preferences, as user studies show). These expectations influence the behaviors that the user might want to see from the animatronics. Due to the layered software architecture, the same conversational agent can control any of our embodiments, without modifications of the state machine. A diversity of embodiments is fully intended, since users may have strong individual preferences for their personal animatronics.

Although the main function of the Intermediary's animatronic device is enhancing communication and alerting, ideally, it is *not* just like any other piece of equipment, and certainly not just like a cellphone. It rather should be regarded as a 'sentient companion' (although not in the literal sense) that keeps the user's company, much like a pet dog or another small, tamed creature. Such a view suggests some of the ways an Intermediary could be embodied—the ways it could look like.

Since the animatronics part of the Intermediary is a personal companion to the user, the metaphors were explored that we are used to when it comes to pet like companions.

The most famous one is probably the parrot sitting on the mystical sailor's shoulder (Figure 7). Another one is the snake wound around the handler's neck. Some metaphors are more contemporary, like a small rodent 'living' in the shoulder/neck area of a punk rocker. The last two mentioned, however, do not guarantee wide public acceptance, because of the ambivalent connotation of snakes and rats, and therefore should probably be avoided.

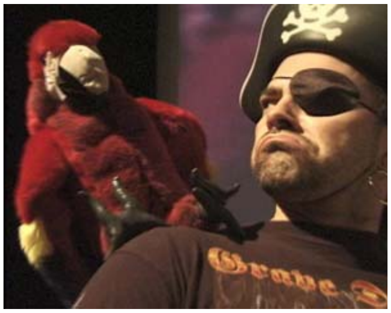


Figure 7: Pirate with parrot

However, there are more ways an Intermediary can be embodied, keeping in mind that one of the most important reasons to embody the Intermediary is *to provide a natural and clear focal point of attention for the people around the user.* In other words: it has to be clearly visible to the people around the user. One such embodiment could be a hamster (or similar sized creature) sitting in the user's chest pocket. This location is highly visible to the people around the user, and includes the important option of looking *up* to the user.

As mentioned earlier, another important reason to embody the Intermediary is to use socially intuitive cues to interrupt and alert, instead of ringing or vibration. One of the strongest social cues is gaze. Therefore, it is important that an Intermediary can look at people, and at the user specifically, with big eyes. As a contrast, the Intermediary could be asleep when not in use. This can include slight breathing movements to make it still appear ‘alive’ (in a wider sense).

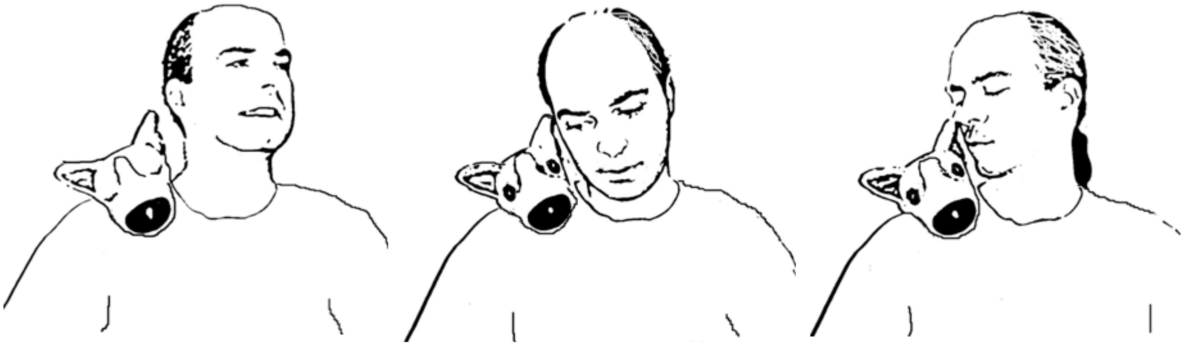
In general, the most generic mapping between the animatronics behaviors and meaning is as follows:

Sleeping, breathing	Idle, nothing important going on
Waking up, looking around, seeking eye contact	Get attention from user and co-located people

In the following, four different generic types of embodiments are presented that differ in their respective functional advantages and disadvantages. Then the three embodiments that were built are described in detail.

4.3.2. Creature resting on shoulder

Features: Opens and closes its big eyes; touch sensitive nose and ears
Advantages: Good visibility to other people; rests easily on shoulder
Disadvantages: Only one degree of freedom (only its eyes are animated)



Sleeping: eyes closed Attention seeking: eyes open Communicating

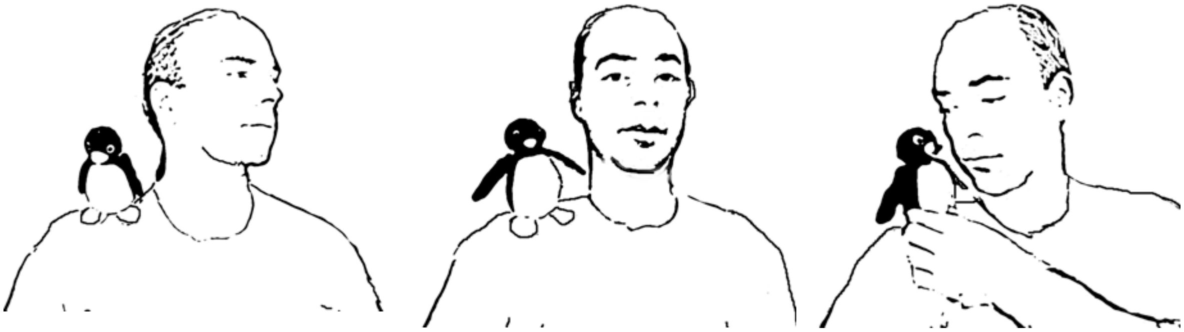
Figure 8: Creature resting on shoulder

Although having a creature resting on a user's shoulder (Figure 8) is highly visible to co-located people (which is the desired effect), the user himself can't see the eyes of the creature if its head is not turning. Therefore, opening its eyes could be accompanied by a very low volume sound, only audible to the user. Such a sound would also mask the sound of the actuators, if they were based on motors and gears. (The masking issue gets irrelevant if quiet actuators are used, such as magnetic actuators or actuators based on shape memory alloys.)

This instantiation is based on a ‘lazy animal’ resting its (oversized) head on the user’s shoulder. A typical example is *TarePanda*™ (a very flat panda stuffed animal), as well as the Artlist International© *THE DOG*, which has an extremely oversized nose and head section. Both of these animals have big eyes, which makes them perfect to grab attention by just opening their eyes. In addition to that, these dolls incorporate all features that seem to influence the ‘cuteness’ of a creature: big eyes, high forehead, big head compared to body, short arms and legs. Cuteness may be important to increase the social acceptance of an Intermediary. In addition, it is often associated with young creatures, like puppies, which are given more freedom in case of misbehavior, since the creature is still in its infancy, and just doesn’t know any better. Therefore, people are more forgiving with interruptions from creatures obviously still “in training.”

4.3.3. Bird standing on shoulder

Features: Moving head up/down, or eyes opening/closing; wings flapping; touch sensitive wings; head turning towards user
Advantages: Very good visibility on shoulder, can talk directly into user's ear
Disadvantages: Difficult to mount/balance on shoulder



Sleeping: looking down

Attention seeking: looking straight, flapping wings

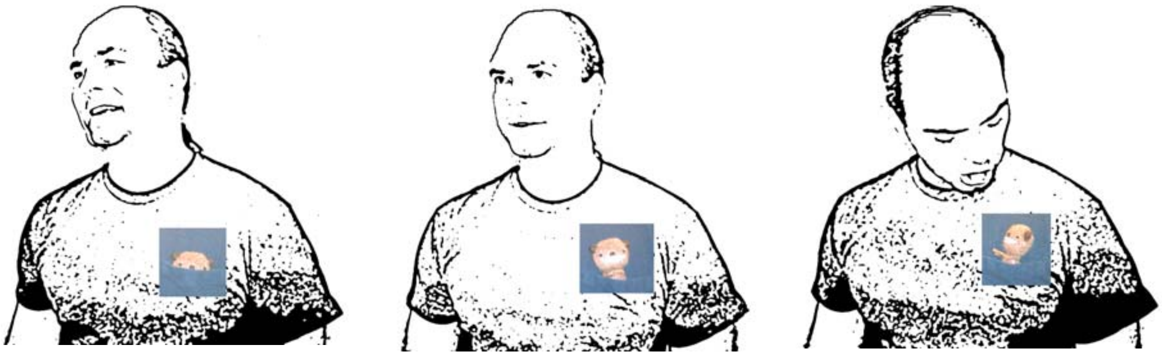
Communicating: head turned sideways

Figure 9: Bird standing on shoulder

Although balancing a bird on one's shoulders (Figure 9) is non-trivial, sitting on the user's shoulders has the obvious advantage of being very close to the user's mouth as well as one of his ears. Because the microphone is close to the user's mouth, his voice is picked up well even if talking in a low volume; and because the speaker is close to the user's ear, especially when the user turns towards the Intermediary, playback volume can be very low and still acceptable for the user.

4.3.4. Creature in chest pocket

Features: Moves in and out of chest pocket (vertically), turns upwards towards user
Advantages: Convenient to carry; small
Disadvantages: Difficult to integrate all elements into a chest pocket sized animal; not as visible as the other instantiations



Sleeping: eyes closed, sitting deep in pocket

Attention seeking: eyes open, looking straight, peeking out of pocket

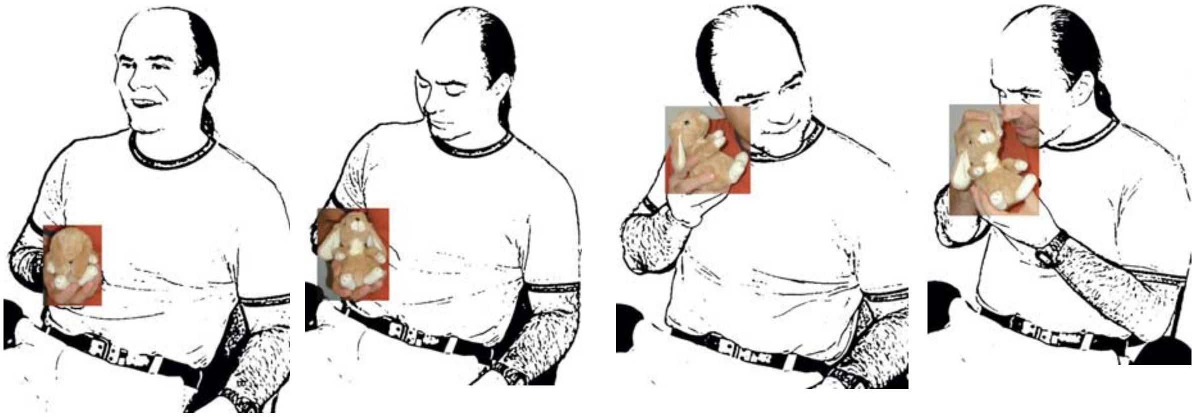
Communicating: head turned upwards

Figure 10: Creature in chest pocket

This instantiation (Figure 10) is inspired by a hamster that sits in the user's shirt pocket, usually asleep, but wakes up when it has to alert, peeks out and looks up to the user when it wants his attention. A possible version would be a Beanie Baby sized doll, or a custom made stuffed animal (like in Figure 10).

4.3.5. Creature in hand and on table

Features: Moving head up/down (big ears covering eyes); touch sensitive ears
Advantages: Doesn't have to be worn, can sit on desk by itself
Disadvantages: Has to be carried around



Sleeping: eyes covered by big ears, looking down

Attention seeking: looking up (uncovering eyes), head bent back

Communicating (talking to user)

Communicating (listening to user)

Figure 11: Creature in hand and on table

As mentioned above, making the creature appear cute is important to increase its social acceptance for co-located people. This specific instantiation (Figure 11) profits from the very cute movement of a small rabbit baby being curled in during sleep, almost spherical in shape, and then stretching its back when waking up. When asleep, its eyes are covered by its floppy ears, but are uncovered in a very cute way when waking up.

This is a typical example of a cute movement, which can be as important as cute static features. It is likely that such movements are slow, never abrupt or fast, and possibly with non-linear acceleration and deceleration.

Since cuteness does not have to coincide with 'life-likeness,' it is possible to explore non-lifelike entities as Intermediaries that become attractive and socially acceptable through their mere movements. The movement of "unfolding" seems a promising candidate. A good example it the so-called robotic calculator that unfolds and stands up, which is an amazingly cute feature since the spring is damped heavily to allow for a very smooth and slow unfolding process. Another possibly cute movement could be a creature coming out of its nest or 'house', like a hermit crab or a turtle peeking out of its shell.

Other possible locations for the embodiment include:

- Hanging in front of chest, with necklace
- Wrapped around neck, as a scarf (octopus, snake)
- Wrapped around upper or lower arm
- On user's back or over shoulder: e.g., a monkey disguised as a backpack or shoulder bag.
  Advantage: enough space for adding sub-systems; can "hold" or "hug" the user naturally
  Disadvantage: much larger than cellphone
- Finger mounted, fingertip mounted (thimble), thumb nail mounted.
  Disadvantage: too small to incorporate all necessary subsystems

Other possible degrees of freedom for the embodiment may include:

- Opening/closing pupils (making big eyes)
- Tilting head sideways (may increase perceived cuteness)
- Wiggling ears or tail
- Raising eyebrows
- Crawling up and down the user's sleeve (attached to lower arm)
- Shrinking shoulders
- Waiving with paws (if sitting in chest pocket)
- Nose movement (sniffing, like Ocha-Ken™)
- Slightly breathing (chest movements)
- Blowing up cheeks (like hamster)
- Moving and glowing up whiskers
- Rattling (snake)
- Moving eyes on eyestalks

Clearly there is a design and fashion aspect to an Intermediary embodiment. Cellphones are becoming fashion statements, a trend that will soon become the main reason to buy new communication devices. Although it will be very difficult to keep up with the quickly changing fashion trends, there are things that would increase the acceptance of an Intermediary to fashion conscious users, e.g., can if it can be worn in more than one location.

4.4. Animatronics

In this section, I will describe the Intermediary embodiments that were developed as part of this thesis work.

4.4.1. Three generations

Several generations of animatronics were developed during the last years. All of them were originally stuffed animals that were heavily "enhanced" and contain some or all of the following subsystems:

- Actuators and sensors
- Wireless transceiver (i.e., Bluetooth for duplex audio and data)
- Audio (audio amplifier, speaker, microphone)
- Animatronics control (converting actuator and sensor signals)
- Batteries and power conditioning
- Skeleton and skin

There are three consecutive generations of animatronics:

- Parrot
- Bunny
- Squirrel

Each has different capabilities, for example, different degrees of freedom and different audio/data links.

Actuation
The parrot has four degrees of freedom: two for the neck (up-down, left-right), and both wings separately. This allows the bird to look up, look around, express different patterns of excitement and frustration with its wings, etc.

Both bunny and squirrel have also four DOF: two for the neck and spine, and both eyelids. The initial posture is curled up; they wake up with an 'unfolding' movement. They then can look around, and together with fine eyelid control express surprise, sleepiness, excitement, etc.

In order to create a realistic eye opening and closing expression, both bunny and squirrel are able to move both upper and lower lids, using small rubber bands as lids that are pulled back simultaneously by a micro servo via thin threads.

All actuators are independent channels that are fully proportional with a resolution of 100 steps from one extreme to the other.

The animatronics do not try to express emotions per se. Since they mainly use gestures and gaze, they do not employ complex facial expressions other than moving eyelids, and have no need for mobility (i.e., no walking).

Wireless link

Although in the future, the animatronics may be controlled directly by the user's cellphone, or the animatronics will contain the cellphone, the current animatronics prototypes are implemented with a 'remote brain' approach: they are computer-remote controlled, but completely wireless and self-contained devices.

The three generations of embodiments differ in their wireless links: the parrot has a simplex data link and no audio capabilities. The bunny sports a simplex data link as well as half-duplex audio. And the final generation, the squirrel, has both full duplex audio and data link.

The parrot and the bunny are controlled via radio control gear that is used by hobbyists to control airplanes and boats. This channel is simplex, with a range up to 100 meters indoors.

The animatronics control software sends outs serial signals over RS232 to a "glue" board containing a microcontroller that generates a transmitter-specific pulse width modulation signal, which is fed into the customized radio transmitter via its 'buddy plug.' The radio receiver in the animatronics receives these commands and moves the servos accordingly. The R/C and animatronics in the parrot (receiver, servos, batteries, mechanics) are off-the-shelf modular components used by hobbyists. The bunny, with its smaller body size, uses much smaller components that are intended specifically for ultra light R/C airplanes.

The second-generation embodiment, the bunny, also contains a half-duplex audio transceiver (FRS radio module in the 462MHz spectrum). Channel control is done via pressing one of the bunny's ears, which contains a switch that triggers the push-to-talk button on the radio ("squeeze-to-talk" metaphor). On the desktop computer side, the push-to-talk button is pressed via yet another microcontroller "glue" board that is connected to the serial port of the PC: whenever the PC wishes to play back audio on the animatronics, the PC can open the channel automatically and play the audio over its soundcard to the animatronics. In a similar way, the PC receives the audio coming from the animatronics via its microphone input, where it gets digitized and further processed.

The most advanced embodiment, the squirrel, sports a fully digital link for both audio and data. On the desktop computer side, a Bluetooth class 1 transceiver is used with modified antenna to achieve a range of 40 meters indoors. On the embodiment side, a Bluetooth class 1 module with a ceramic antenna is used. This Bluetooth link allows simultaneous duplex audio and duplex data transmission, and replaces the bulky R/C transmitter and half-duplex radio of our earlier prototypes. The duplex audio capability enables not only asynchronous voice instant messages between caller and user, but also a full duplex phone conversation. The duplex data channel allows sending back sensor data from the embodiment to the animatronics control software.

Figure 12 shows a summary of the differences of the three generations.

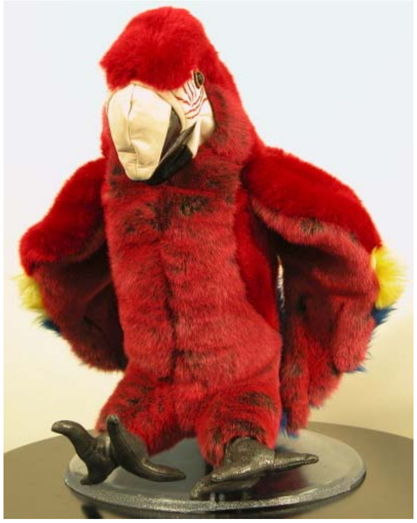
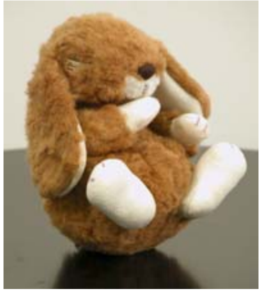
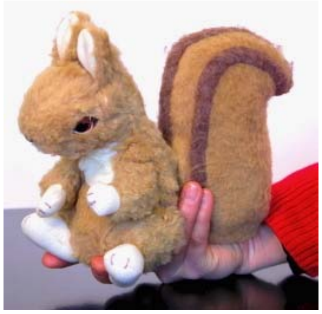
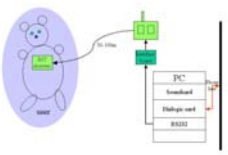
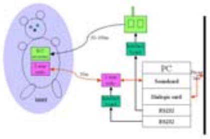
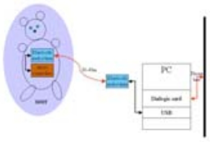
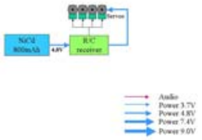
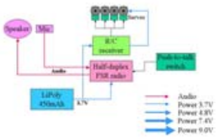
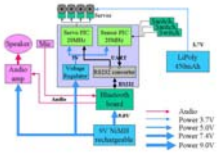
			
Size	38 cm tall	11 cm tall	12 cm tall
Data link	Simplex analog	Simplex analog	Duplex digital
Audio link	N/A	Half-duplex analog	Duplex digital
DOFs	Neck (2), wings (2)	Neck (2), eyes (2)	Neck (2), eyes (2)
Remote communications			
Animatronics control			

Figure 12: Three generations of animatronics. The animatronics control and remote communications diagrams will be explained in detail in the respective sections.

In the following sections, I will describe the three generations of animatronics in detail.

4.4.2. Parrot

The parrot animatronics is based on a beautiful scarlet macaw hand puppet[3]. This puppet was ideal since it was already empty inside (unlike real stuffed animals), but still had to be modified heavily.

[3] http://www.puppetworld.net/

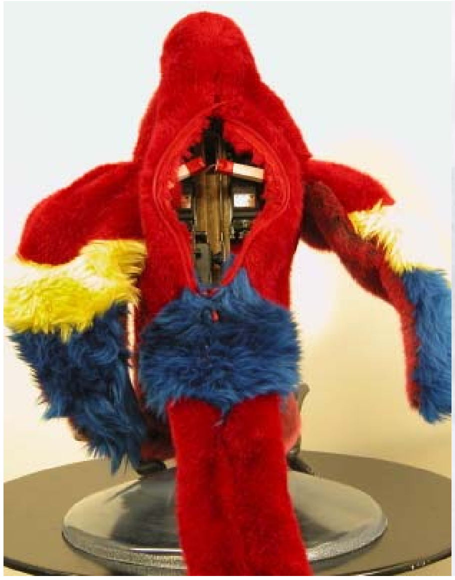


Figure 13: Parrot with open back

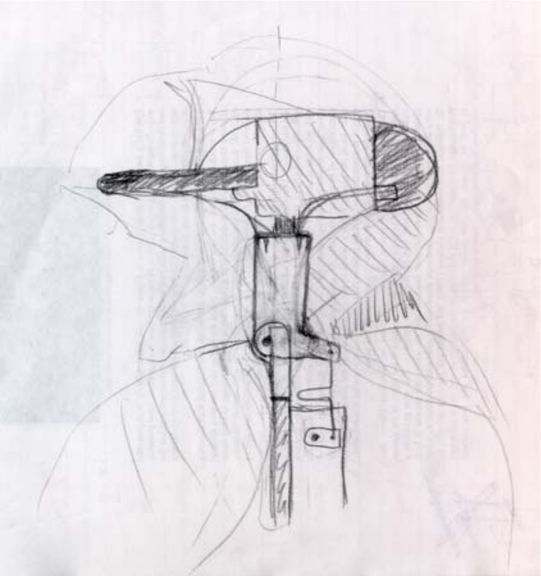


Figure 14: Early design sketch

Mechanics

A zipper was inserted into the back that allows convenient access the interior of the bird (Figure 13). The content of the head was emptied to accommodate the neck servos (Figure 14).

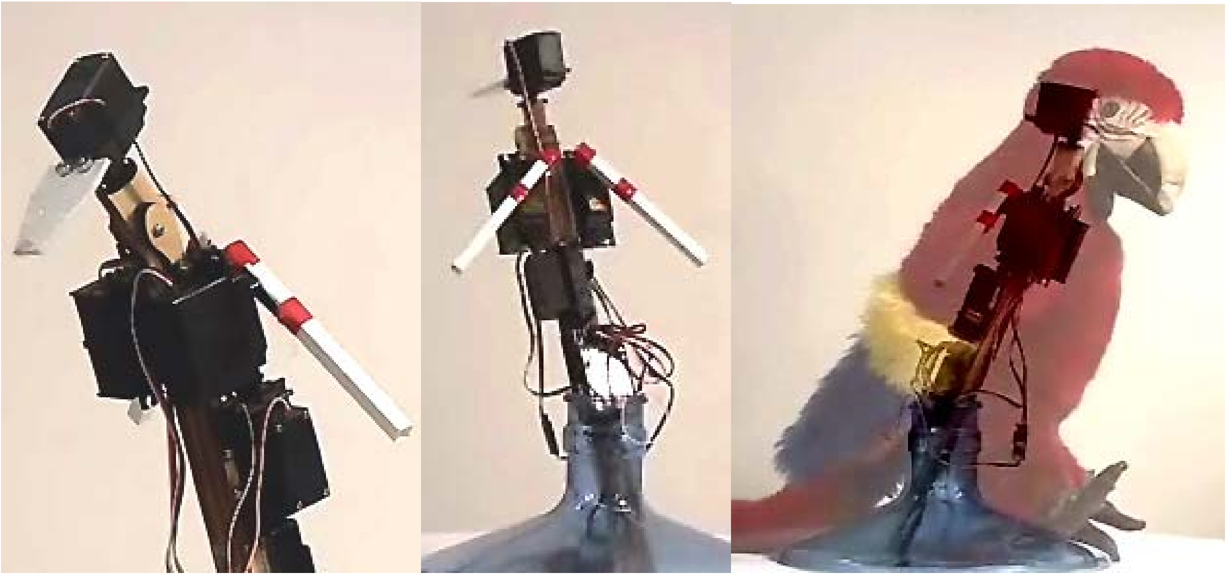


Figure 15: Parrot animatronics

The parrot has four degrees of freedom: two for the neck (up-down, left-right), and both wings separately. This allows the bird to look up, look around, and express different patterns of excitement and frustration with its wings.

The neck consists of a servo that can turn the head sideways. This servo is attached to the spine with a 'nodding' joint. A second servo moves the whole first servo forward and backward (nodding motion) via pushrod and clevises.

The wing servos are attached on the side of the spine, and a square plastic tube extends the servo horns into the wings (Figure 15).

Remote communications
Figure 16 shows the communication architecture of the parrot.

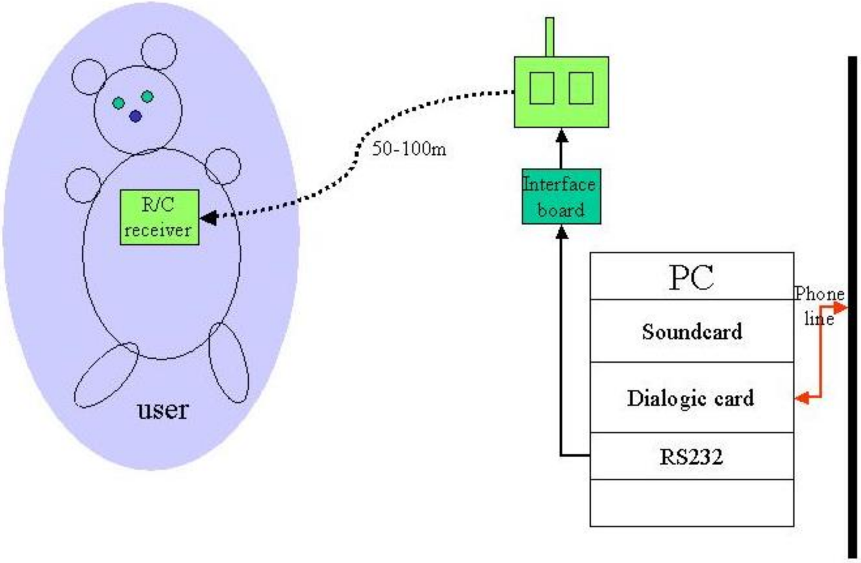


Figure 16: Communications for the parrot

The animatronics sequencer and server (section 4.5) running on the remote PC sends outs serial signals over RS232 to a “glue” board. This board contains a microcontroller that generates a pulse width modulation signal sequence. This signal is fed into the customized radio transmitter (Futaba T6XA) via its ‘buddy plug.’ The transceiver sends this PWM train signal over radio (72MHz spectrum) to the receiver in the animatronics, where it moves the servos accordingly. The R/C and animatronics in the parrot (receiver, servos, batteries, mechanics) are off-the-shelf modular components used by hobbyists for model airplanes, cars, and boats.

This communication solution has the advantage of a good radio range: outdoors it is up to 300 meters, indoors about 100 meters. This type of communication is stable, since the components are commercially available and well developed. However, it is a simplex link, so the data flows only in one direction, from transmitter to receiver.

Furthermore, such R/C transmitters are built for a human operator, and therefore typically do not have a digital control interface. There are only few attempts to interface an analog R/C transmitter with computers in order to give the software full control over the transmitter’s functionality. Therefore, means had to be developed to allow the computer to control the transmitter.

R/C transmitter modifications

Most commercially available R/C transmitters have a so-called "buddy plug." A proprietary cord is plugged into two transmitters and connects them. One transmitter is operated by a less experienced user, the other by an expert or teacher. It allows the teacher to override the commands of the student with the flick of a switch in case of catastrophic pilot errors.

This buddy plug is the only directly available way to feed a control signal into a transmitter. The plug accepts an analog signal, a transmitter-specific PWM pulse train: it varies in terms of the amount of channels the transmitter has, as well as other parameters. Generally, servos are controlled by a pulse of variable width. The angle of the servo arm is determined by the duration of a pulse. The servo expects a pulse every 20 milliseconds. The width of the pulse will determine how far the motor turns. A 1.5-millisecond pulse, for example, will make the motor turn to the 90-degree position (often called the neutral position). If the pulse is 1.0 ms, then the motor will turn the shaft to closer to 0 degrees. If the pulse is 2.0ms, the shaft turns closer to 180 degrees (Figure 17).

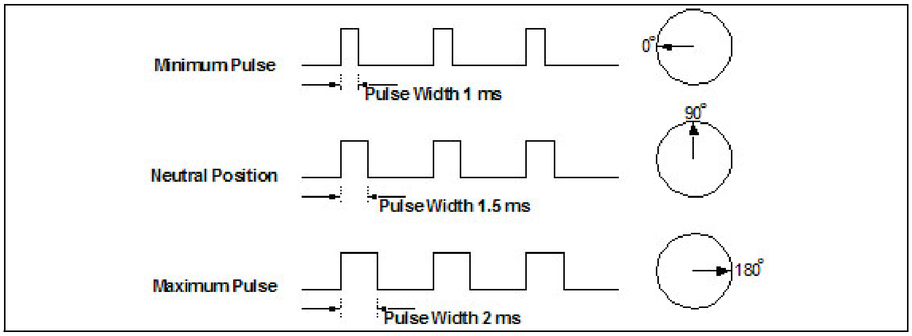


Figure 17: Servo pulse width modulation

The transmitter needs one pulse for each servo (6 in our case), and has to repeat the complete pulse train every 20ms in order to maintain stable communication with the receiver. A PIC microcontroller (16F84A) is used that receives servo commands from the PC via RS232, and generates the continuous PWM signal for the transmitter.

Animatronics control

Because of the commercially available and well-developed radio gear, the animatronics control within the parrot is rather simple (Figure 18). The receiver gets the signals from the transmitter, and distributes them to the servos. A single 800mAh Nickel Cadmium (NiCd) battery powers both receiver and all servos.

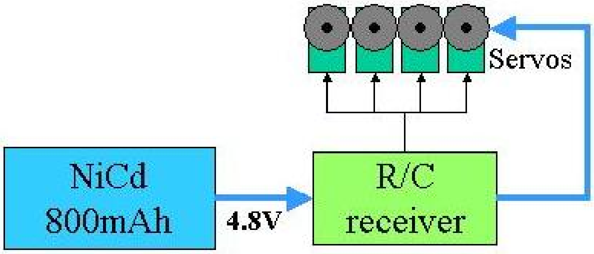


Figure 18: Parrot animatronics control

The parrot was used to explore the animatronics and remote communication infrastructure, but since it lacked audio capabilities, it never became a full Intermediary.

However, it was used in a demonstration of the concept of an Intermediary during a presentation. The bird was mounted on the shoulder of a pirate, who 'interrupted' a talk—which incidentally was about embodying agents into stuffed animals. During the show, a confederate in the audience remotely controlled the parrot (thanks to David Spectre). The life-likeness of the animatronics was quite stunning, as documented in a short video[4] (see also Figure 19).

Figure 19: Parrot show

4.4.3. Bunny

The next generation animatronics was built on top of a cute stuffed animal in the shape of a bunny, about 11cm tall (Figure 20). The bunny was chosen specifically for its cuteness, but also because of its size: although it fits perfectly into a hand, it has enough space inside to accommodate all electronics and mechanics.

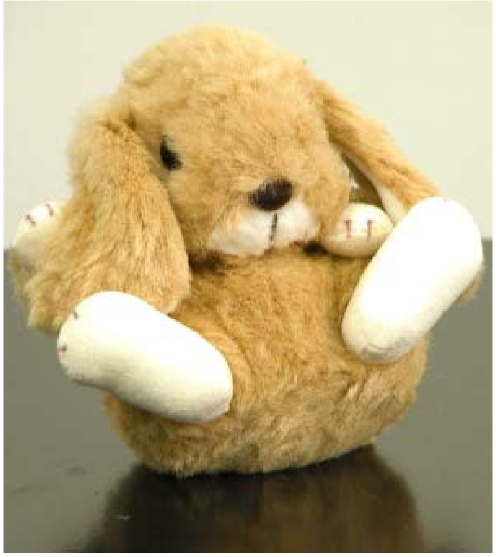


Figure 20: Original Starchild© bunny

[4] http://web.media.mit.edu/~stefanm/phd/videos/

As a stuffed animal, its basic posture is curled up, almost spherical in shape. In this position, the floppy ears tend to cover the eyes. If the bunny raises its head, the ears uncover the eyes.

In order to fit in all the components, all the stuffing was removed, and the seam on its back opened and replaced with a zipper.

Mechanics

The neck consists of two servos (Cirrus CS-6.2) that are connected head to head, with an angular offset of 90 degrees (Figure 22).

This neck construction allows the bunny to look left and right with a 90-degree angle, and independently raise its head with about the same angle.

The neck and spine are made of a half-inch wide strip of brass. Both servo arms are screwed onto an L shaped connector with a 90-degree offset. The lower servo is screwed to the brass spine that also serves as a base. Because the bunny would not stand by itself on this base, the metal lid of a glass jar was added. Since the base was still not heavy enough, the lid was filled with a dozen quarters that were taped together.

Instead of actuating the paws, it was decided to make the eyes open and close. The eyes of the robotic cat Necoro (by Omron™) were inspiration for a solution that can move both upper and lower lids.

However, the head and the eyes of the bunny were much smaller, which was posing a problem for the actuators.

Two micro servos (Cirrus CS-4.4) were found that fit in the bunny's head. A mechanism was developed that allows moving both upper and lower lids using small rubber bands. Small rubber bands were slit in the middle and wrapped tightly around commercially available Teddy Bear eyes. The lids are pulled back by the micro servo via thin threads (Figure 21). Both servos were taped to a Balsa head plate with several stabilizing Balsa elements. The head plate itself was attached to the upper neck servo (the one that make the head turn left and right).

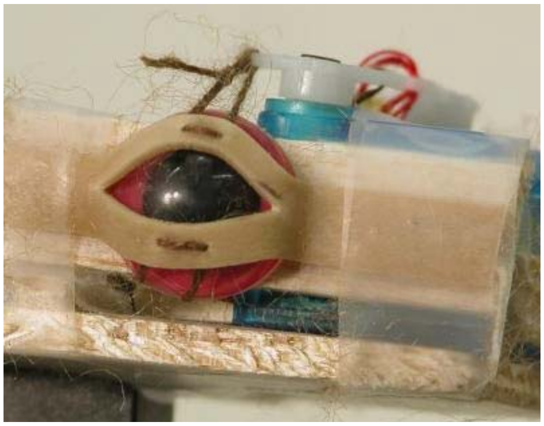


Figure 21: Bunny eyelids

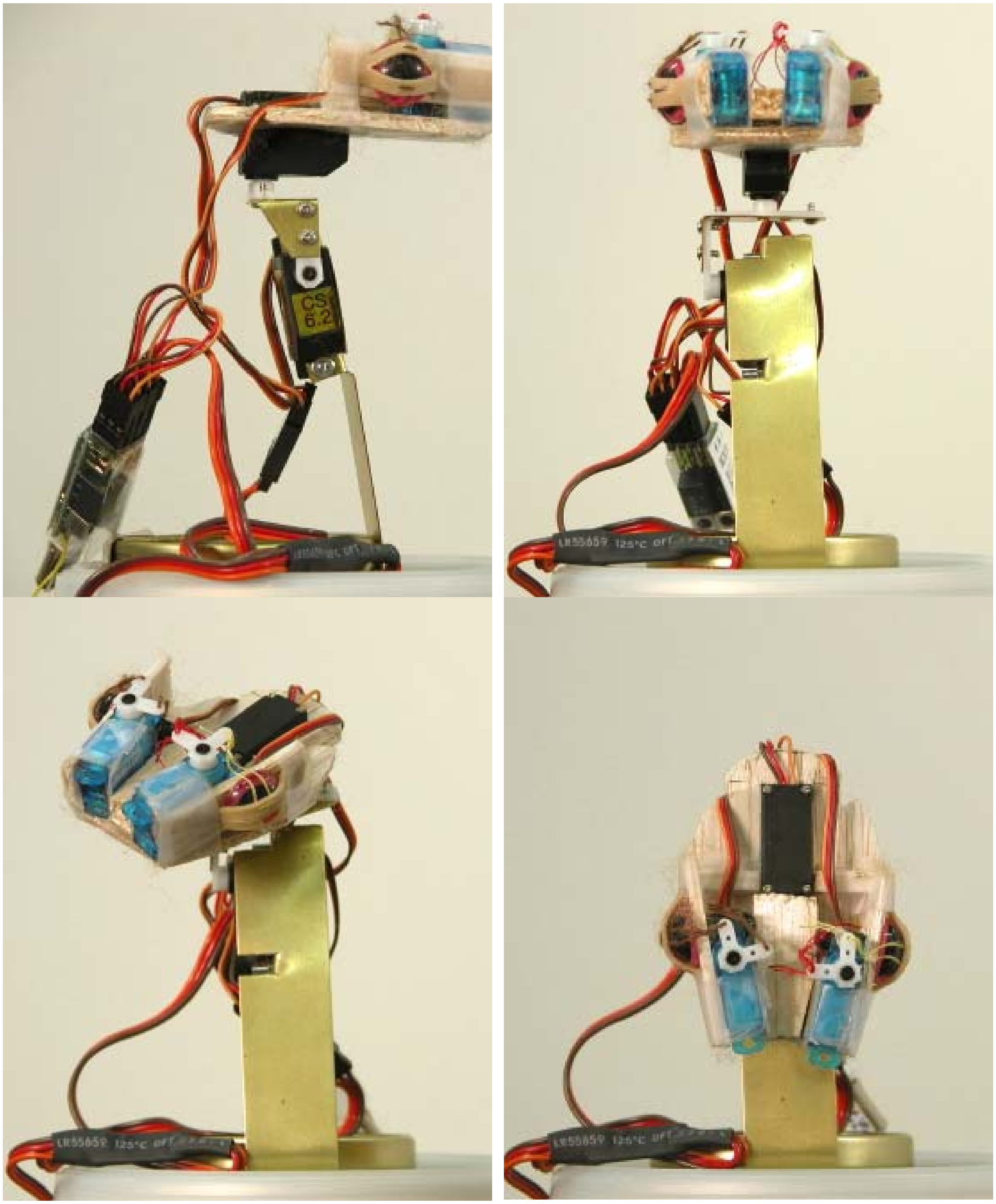


Figure 22: Bunny skeleton and neck; in fully upright posture, it stands 11cm tall.

This construction enabled a very life-like movement of the eyelids. However, the rubber bands become brittle over time and have to be replaced once in a while. Furthermore, the knots of the thin threads that open the lids tend to come loose after some time, and have to be re-fastened. In order to make the lids slide nicely back to closed position (there is no force that closes the lids other than the inherent spring force of the rubber), olive oil as 'eye drops' was used. The oil also keeps the rubber bands from drying out and becoming brittle fast.

Although the skeleton of the bunny may not look life-like at all, the bunny with zipped up skin is adorable (Figure 23).

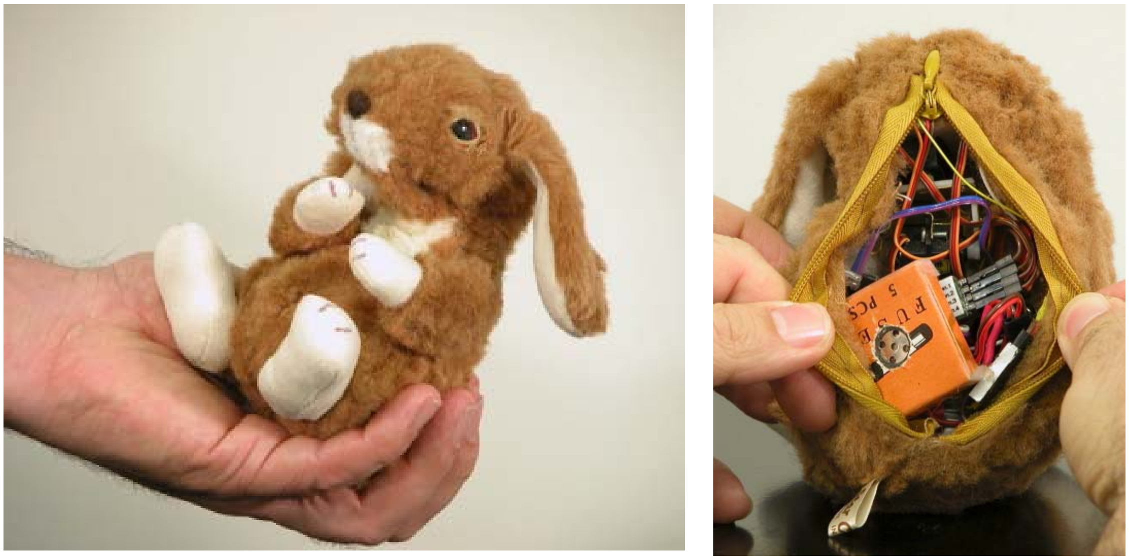


Figure 23: Bunny looking up (left), and with open back (right)

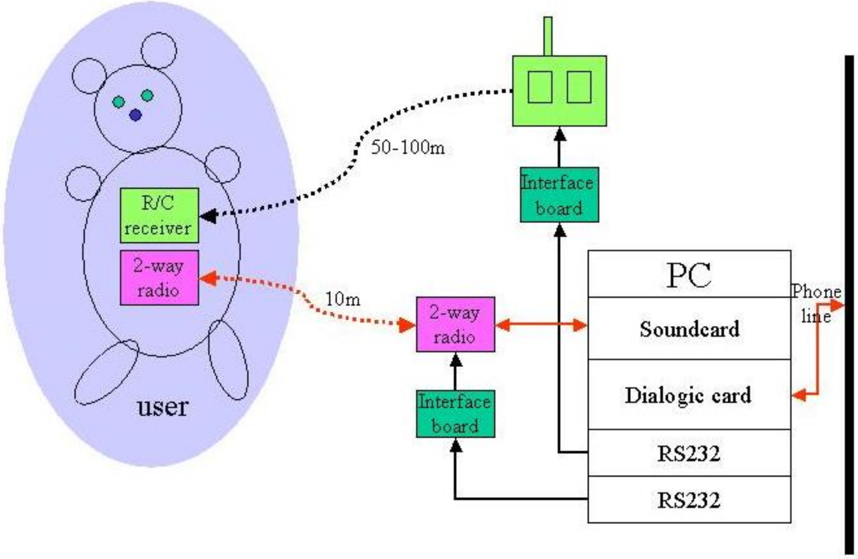


Figure 24: Communications for the bunny

Remote communications
Figure 24 shows the remote communications architecture of the bunny. It uses a similar configuration as the parrot, based on commercially available R/C transmitter, interface board, and R/C receiver.

The half-duplex audio link between animatronics and computer is new. Main goal of this prototype was to demonstrate the Intermediary's voice instant message passing capabilities; therefore, the missing duplex capability was not relevant.

This implementation of an audio link consists of a half-duplex audio transceiver, and FRS radio module in the 462MHz spectrum. Channel control is done via pressing one of the bunny's ears, which contains a switch that triggers the push-to-talk button on the radio. On the desktop computer side, the push-to-talk button is pressed via yet another microcontroller "glue" board that is connected to the serial port of the PC: whenever the PC wishes to play back audio on the animatronics, the PC can open the channel automatically and play the audio over its soundcard to the animatronics. In a similar way, the PC receives the audio coming from the animatronics via its microphone input, where it gets digitized and further processed.

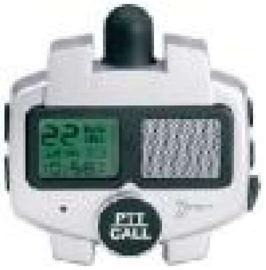


Figure 25: Xact© M2X radio

On the computer side, the transceiver is a Xact M2X (Figure 25). It was modified to accept power by a power supply, bypassing the internal batteries. Furthermore, wires were soldered to the internal push-to-talk button in order to enable external computerized switching (Figure 26).

It took a lot of experimentation until the connection between the radio module and computer became functional. The radio has a headset connection (stereo mini jack), but when connecting the three wires to line in and line out of the sound card with a Y cable, the transceiver goes into transmit (TX) mode even without any signal present at the sound card output. This problem was solved with a component that bridges high and low impedance lines ("direct injection box"). Interestingly, laptop sound card connectors have appropriate impedance so the radio could be connected directly. Since grounding and other noise problems were excluded, the electric characteristics of desktop and laptop soundcard turned out to be significantly different.

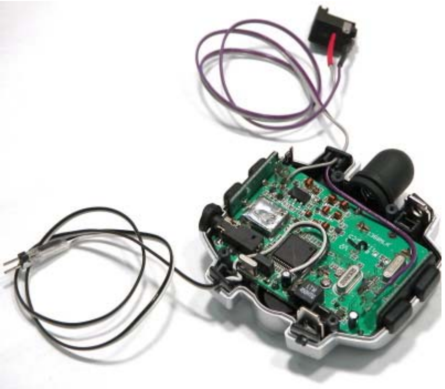


Figure 26: Modified base station transceiver

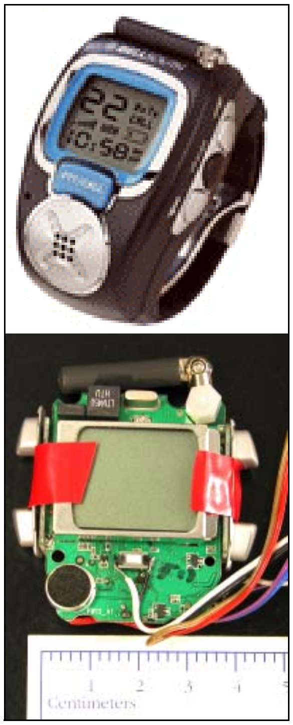


Figure 27: BellSouth™ FW13ZHS radio

On the bunny side, a wristwatch sized transceiver, a BellSouth™ FW13ZHS using the same spectrum, was stripped off its housing to minimize its footprint (Figure 27).

Both the push-to-talk button and the "On" button were bypassed with wires in order to connect external switches. An external push-to-talk button (momentary switch) was put in the right ear of the bunny, allowing the user to grab the bunny's ear when she wants to talk ("squeeze-ear-to-talk" metaphor). An additional momentary switch was hidden in the right foot of the bunny. This allows the user to turn on and off the transceiver without opening the animatronics.

The lithium polymer batteries of the BellSouth transceiver were used to power the whole bunny, including the receiver and servos.

Push-to-Talk control

In order for the computer to be able to press the Push-to-Talk button of the radio transceiver, a PIC microcontroller (16F84A) is used, mounted on an iRX "glue" board, to receive commands from the PC via RS232. One of the pins of the microcontroller is connected to a transistor that in turn is connected to the push-to-talk button of the radio transceiver connected to the computer.

The protocol is as follows: If the PC sends an ASCII "1" the push-to-talk button gets pressed. If the PC sends an ASCII "0" the push-to-talk button gets released again.

In the conversational agent code, a wrapper class was written so that whenever the agent chooses to play back a file on the radio, it would first activate the Push-to-Talk button, and after the sound file playback was finished, release the button via RS232 signals.

Animatronics control

Animatronics control inside the bunny is done in a similar way as in the parrot. Instead of normal sized radio gear, a commercially available micro receiver (Cirrus MRX-4 II 4ch) is used. It gets the signals from the same transmitter (Futaba T6XA), and distributes them to the servos. A single 450mAh lithium polymer battery powers receiver, servos, as well as the audio transceiver (Figure 28).

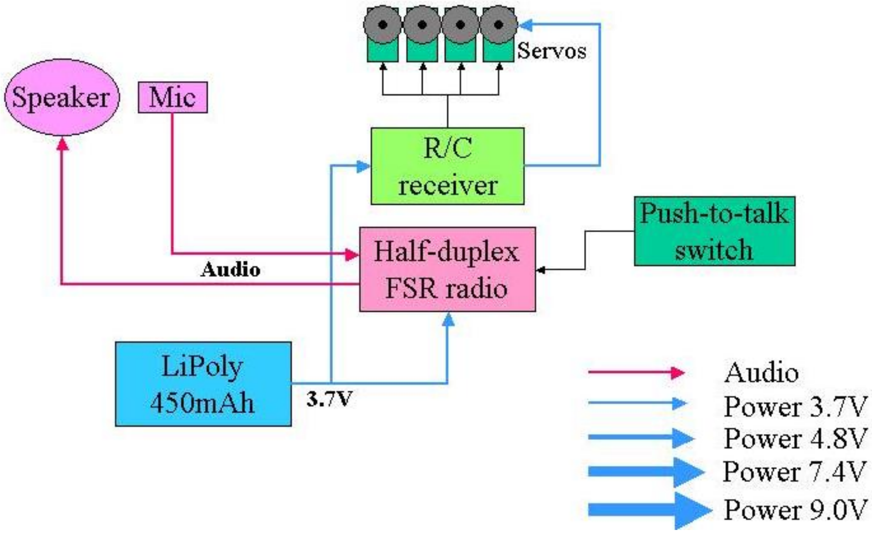


Figure 28: Bunny animatronics control

The micro receiver (Figure 29), used for ultra light airplanes, has the following specifications:

- Dimensions: 10 x 32 x 12mm
- Weight: 9 grams (with oscillator)
- Channels: 4
- Range: 500 meters
- Modulation: FM
- Power: 3.5 – 7V
- Tuner: single conversion, narrow band

Figure 29: Micro receiver

Radioserver extension

A problem with the communication architecture of the bunny is that the range of the audio link limits the range of the animatronics severely. Due to interference, probably from the other transceivers and electronics within the bunny, the audio link would decrease in signal significantly a few meters away from the computer-connected audio transceiver.

In order to circumvent this problem, a server process was created that can run on a remote laptop. It encapsulates the audio functionality of the original conversational agent software, and allows it to run on a remote machine, interacting with the main code via socket messages and shared file systems.

Figure 30 shows an architectural overview of the system that was enhanced with the Radioserver.

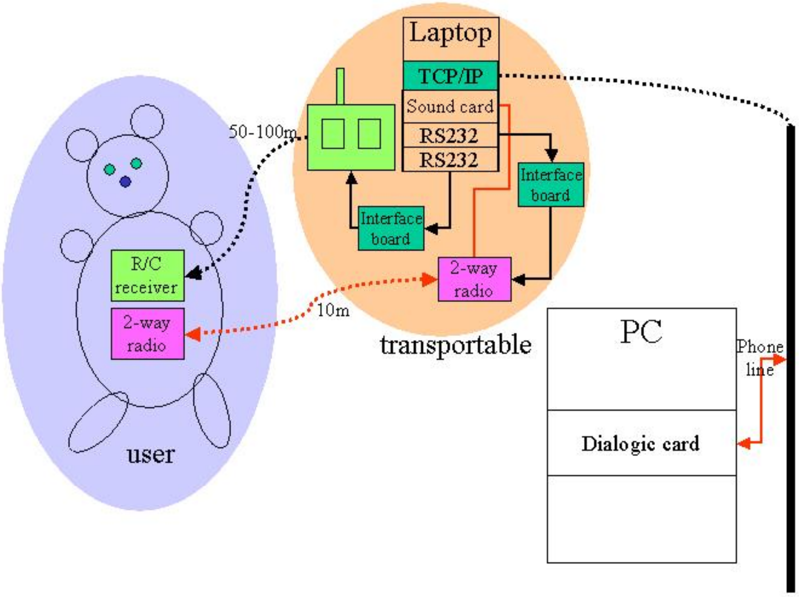


Figure 30: Communications for the bunny, using transportable basestation that runs Radioserver

Due to the Radioserver, this system can be used in any place where WiFi or Ethernet coverage is available. Although there are a number of components that have to be carried around, all of them—including a laptop—fit into a medium sized plastic box. This architecture extends the bunny's 'demoing' range significantly.

The Radioserver has another functionality that was added later. Whenever the user presses the talk button—meaning, squeezing the ear of the bunny—or more precisely, releasing this button, he generates a short noise burst, which is typical for walkie-talkies.

This 'bug' was converted to a feature, since the Radioserver that monitors the audio coming from the transceiver can detect such a noise burst. Such a user button press can be interpreted as positive confirmation signal, or any kind of signal depending on the context. Therefore, the Radioserver is monitoring the audio channel continuously for such clicking sounds, and sends a signal to the main agent code when it detects one.

4.4.4. Squirrel

The squirrel is the most advanced animatronics implementation of the three generations with its Bluetooth duplex audio and data connection. It is based on the same bunny stuffed animal, but its body was modified heavily: the ears were shortened, and a tail was added (Figure 31).

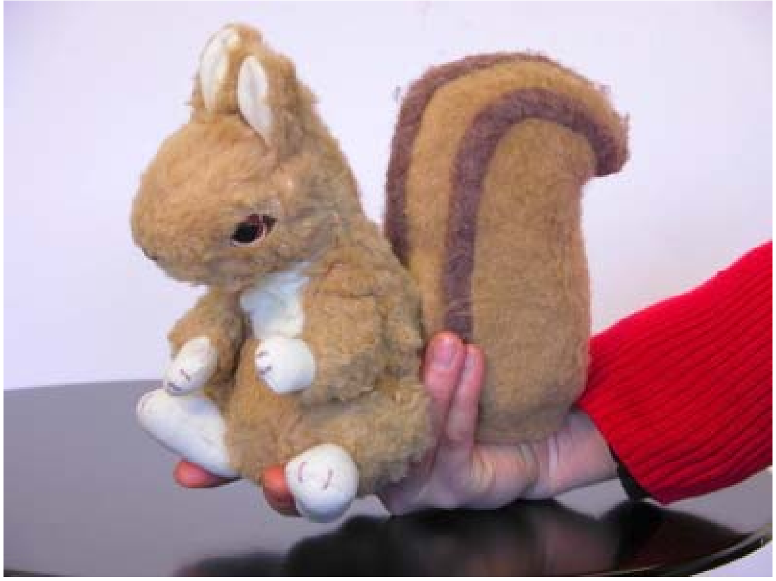


Figure 31: Squirrel

Mechanics
The mechanics are the same as in the bunny. It uses the same skeleton and servos (Figure 32).

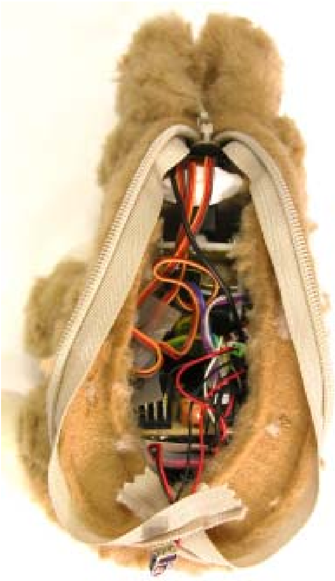


Figure 32: Squirrel with open back

Remote communications
The remote communication architecture is simplified compared to the bunny architecture (Figure 33). This is due to the unified data and audio link, provided by the Bluetooth channel.

On the desktop computer side, a Bluetooth class 1 transceiver (Linksys© USBBT100) is used with modified antenna (2.4 GHz Range Extender) to achieve a range of 40 meters indoors. On the embodiment side, a Bluetooth class 1 module with a ceramic antenna is used. This Bluetooth link allows simultaneous duplex audio and duplex data transmission, and replaces the bulky R/C transmitter and half-duplex radio of our earlier prototypes. The duplex audio capability enables to not only pass asynchronous voice instant messages between caller and user, but also switch to a full duplex phone conversation. The duplex data channel allows sending back sensor data from the embodiment to the animatronics control software.

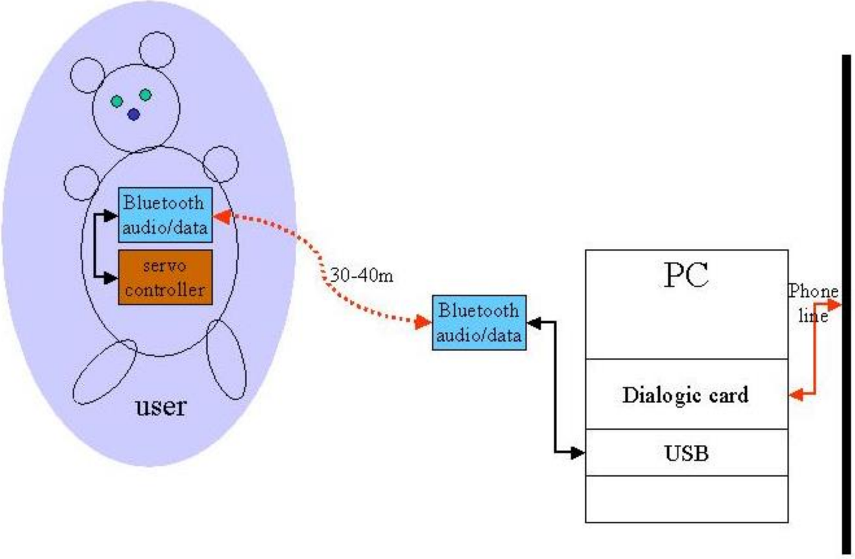


Figure 33: Communications for the squirrel

Animatronics control

Although the communication architecture was simplified compared to earlier prototypes, designing the internals of the Bluetooth animatronics was more complex than with earlier embodiments, and was characterized by many iterations and unsuccessful trials. Without going into details of its development, Figure 34 illustrates these earlier ideas with thumbnails of some of the unsuccessful designs.

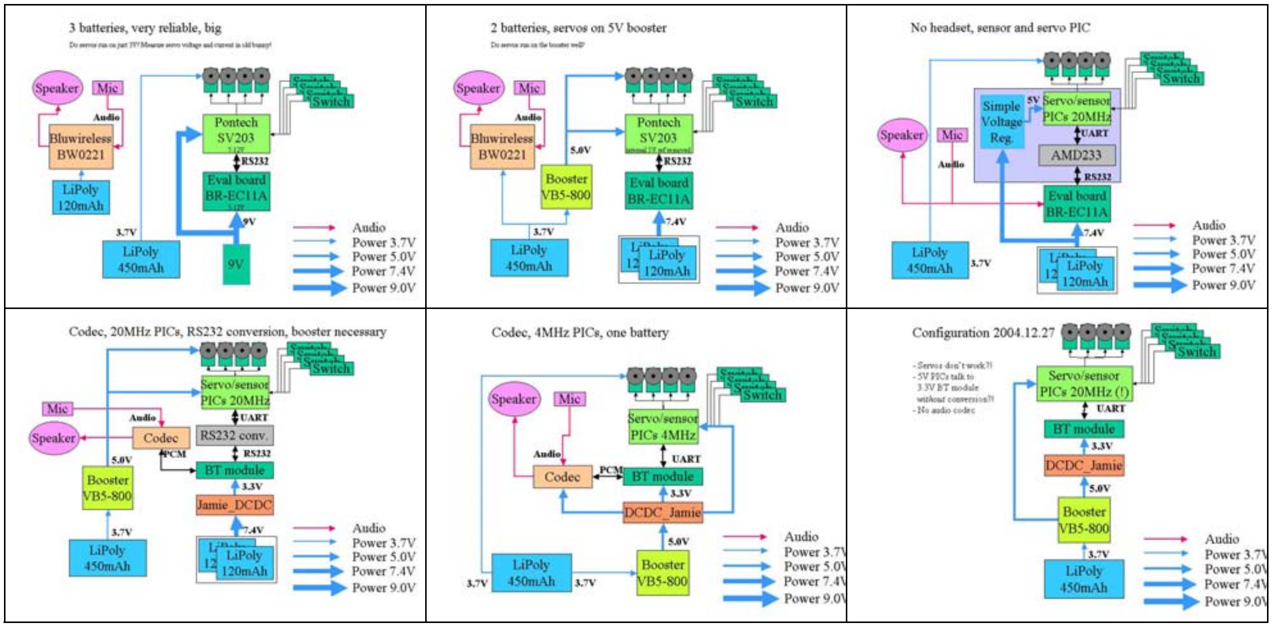


Figure 34: Planning the Bluetooth squirrel architecture

The final squirrel animatronics control architecture is shown in Figure 35, and consists of the following elements:

- Bluetooth board (BlueRadios© BR-EC11A), with onboard audio codec and RS232 UART
- Two microcontrollers (16F87A), one each for servo control and sensor control
- Power conditioning
- Audio amplifier (1 watt)
- Speaker and microphone
- Servos and switches
- Batteries (9V, 3.7V lithium polymer)

Instead of a PCB, all electronic components are soldered to a fiberglass perforation board, a method often used for prototyping where solderless breadboards are too big. All connections between the components are made via thin wires.

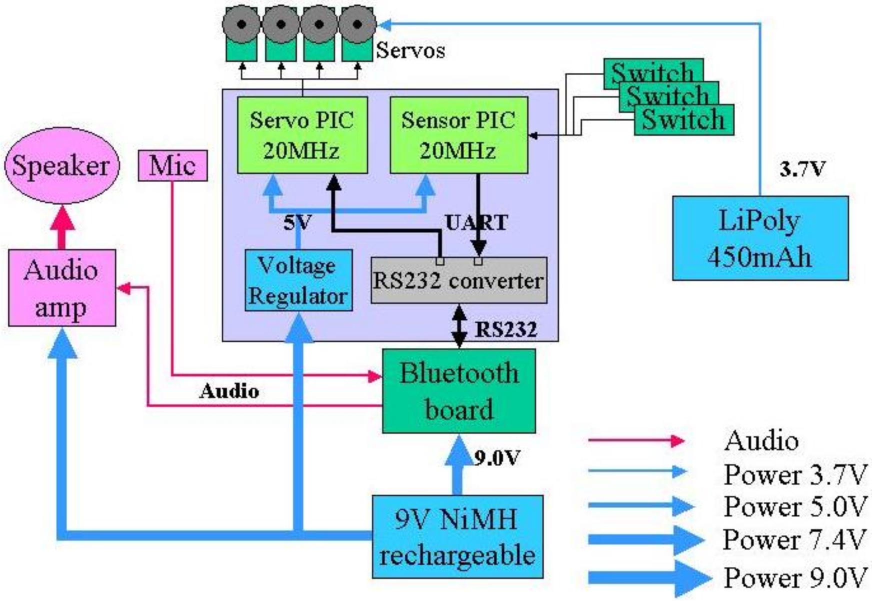


Figure 35: Squirrel animatronics control

In Figure 36, the basic components of the squirrel animatronics control are depicted.

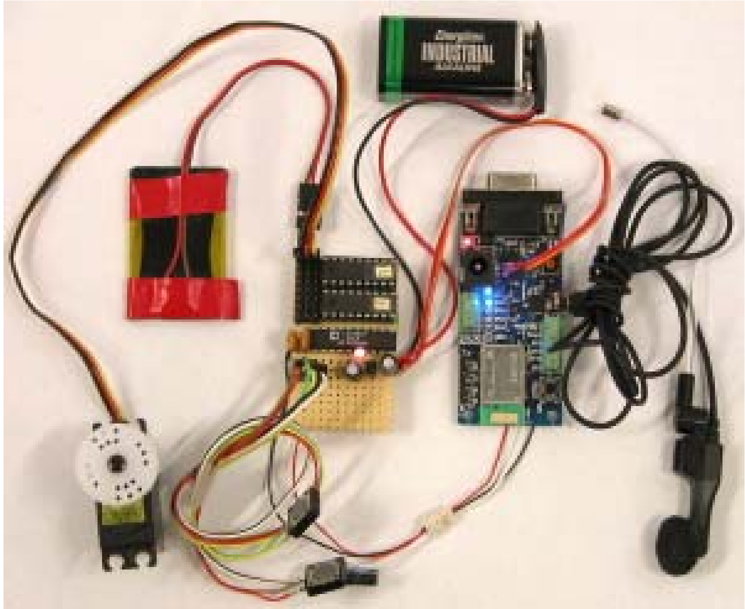


Figure 36: Basic elements of squirrel animatronics control: from top left, clockwise: lithium polymer battery, 9V battery, headset, Bluetooth board, controller board, switches, servo

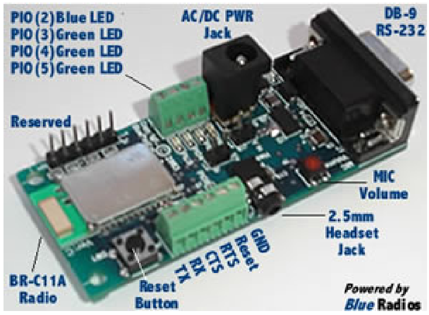


Figure 37: Bluetooth board by BlueRadios©

Bluetooth module
Core of the most advanced Intermediary generation is the Bluetooth transceiver. It is a commercially available board (BlueRadios© BR-EC11A, Figure 37) made for evaluating Bluetooth modules, and comes with a codec, connectors for microphone and line out, UART and RS232 connectors, some programmable status LEDs, a stable power supply, and as well a host of other connectors.

This board is configured and controlled through simple ASCII strings over the Bluetooth RF link or directly through the hardware serial UART. A variety of parameters can be set: some are permanent; some have to be reprogrammed after rebooting.

In order to ‘coerce’ the board into a simultaneous audio and data mode, a sequence of AT commands has to be sent to it upon startup. A microcontroller in the animatronics is used to send the necessary commands at boot time. The same microcontroller is later used to send serial signals back to the dongle connected to the desktop computer.

Controller board
The controller board inside the animatronics consists of two microcontrollers (PIC 16F84A), a RS232 UART converter, power conditioning, LED, some capacitors, two 20Mhz oscillators, headers, and connectors (Figure 38).

The first microcontroller is generating the servo signals from the serial signals it gets via Bluetooth board. The second microcontroller reads the position of all switches and sends back serial signals via Bluetooth board. On one side, the controller board connects to the serial lines of the Bluetooth board. One the other, it houses the connectors for the servos and the switches.

The servo microcontroller, a PIC 16F84A running with 20 MHz, communicates via a 38.4kbps serial interface, and generates PWM signals for 12 servos in parallel with a resolution of 240 steps over 90 degrees rotation. The commands are 2 bytes per servo, one for the ID of the servo, one for the desired position.

The sensor microcontroller, also a PIC 16F84A running with 20MHz, reads the switch positions and sends back serial signals over the Bluetooth connection to the animatronics server. As mentioned earlier, it is assigned another job: at boot time, it first goes through a sequence of precisely timed commands that it sends to the Bluetooth board. After this sequence, it starts reading the position of the switches and sends serial signals back.

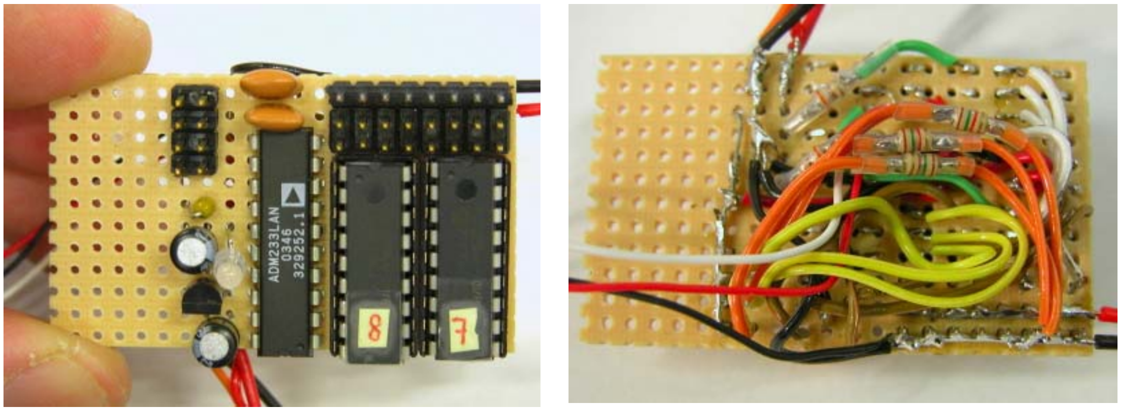


Figure 38: Controller board: top view (left), and back view (right)

Microphone, speaker and amplifier

Although the Bluetooth board has an onboard codec and features a headset output, its audio signal is not strong enough to power a speaker. Therefore, the line out signal is fed into a small 1-watt audio amplifier, which is commercially available as kit (Figure 39).

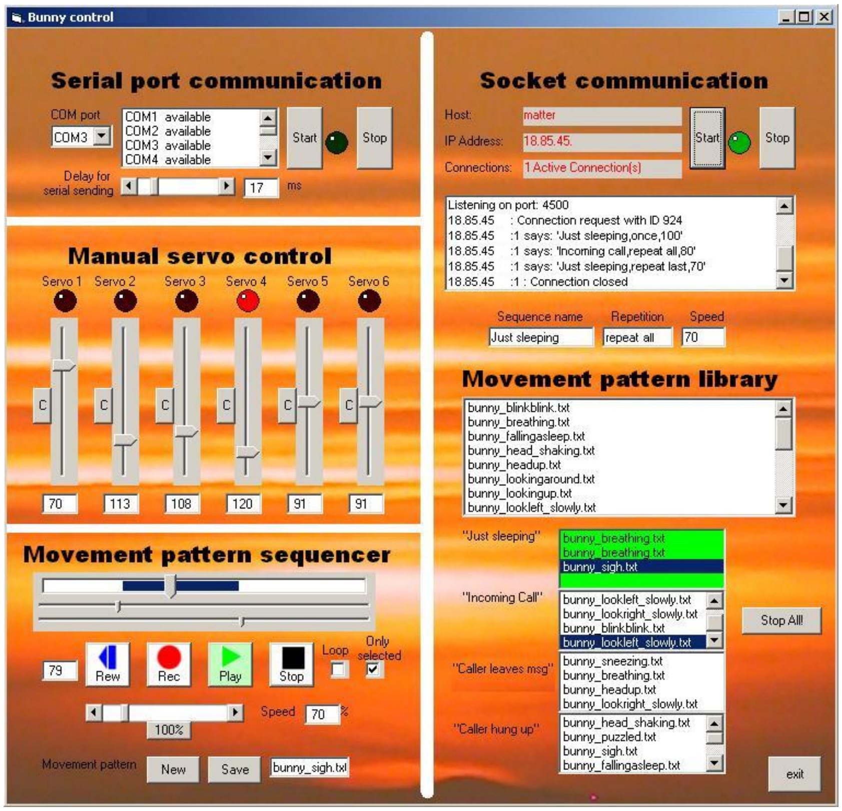
Figure 39: Audio amplifier kit

The output of the amplifier is powering a tiny speaker; both the speaker and the amplifier, together with the batteries, are conveniently located in the bushy tail of the squirrel.

The stereo mini jack connector of a small cellphone headset is plugged directly into the audio connector of the Bluetooth board. The headset's microphone is stripped off of all housing, and the headphone is cut off and replaced with a connector that plugs into the audio input of the audio amplifier.

4.5. Animatronics server and sequencer

All embodiments are controlled remotely by the animatronics server and sequencer (Figure 40). This software serves both as an *authoring tool* to create low and high-level behaviors, as well as *hub* that translates high-level commands from the agent to low-level control signals for the embodiment's actuators, and transmits sensor signals from the embodiment back to the agent.



COM port: COM3

COM1 available
COM2 available
COM3 available
COM4 available

Delay for serial sending: 17 ms

Servo 1: 70 · Servo 2: 113 · Servo 3: 108 · Servo 4: 120 · Servo 5: 91 · Servo 6: 91

79 · 100% · Speed: 70 % · Only selected ☑ · Loop □ · Movement pattern: bunny_sigh.txt

Host: matter
IP Address: 18.85.45.
Connections: 1 Active Connection(s)

```
Listening on port: 4500
18.85.45    : Connection request with ID 924
18.85.45    :1 says: 'Just sleeping,once,100'
18.85.45    :1 says: 'Incoming call,repeat all,80'
18.85.45    :1 says: 'Just sleeping,repeat last,70'
18.85.45    :1 : Connection closed
```

Sequence name: Just sleeping · Repetition: repeat all · Speed: 70

Movement pattern library:
bunny_blinkblink.txt
bunny_breathing.txt
bunny_fallingasleep.txt
bunny_head_shaking.txt
bunny_headup.txt
bunny_lookingaround.txt
bunny_lookingup.txt
bunny_lookleft_slowly.txt

"Just sleeping": bunny_breathing.txt, bunny_breathing.txt, bunny_sigh.txt

"Incoming Call": bunny_lookleft_slowly.txt, bunny_lookright_slowly.txt, bunny_blinkblink.txt, bunny_lookleft_slowly.txt

"Caller leaves msg": bunny_sneezing.txt, bunny_breathing.txt, bunny_headup.txt, bunny_lookright_slowly.txt

"Caller hung up": bunny_head_shaking.txt, bunny_puzzled.txt, bunny_sigh.txt, bunny_fallingasleep.txt

Figure 40: Screenshot of parts of the animatronics sequencer and server

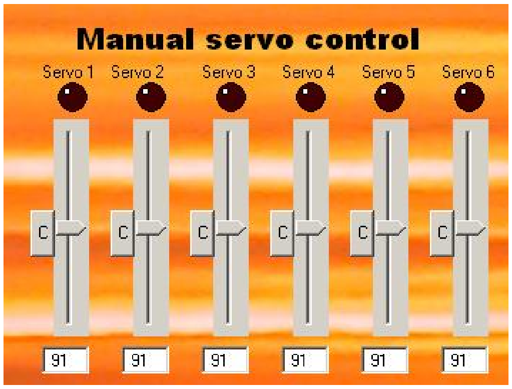
In the future, the software with hub functionality may run directly on the user's phone, whereas the authoring tool may remain on a desktop.

The animatronics server and sequencer incorporates the following functionality:

- Record and modify behavior primitives in loops
- Compose primitives into behavior sequences
- Map behavior sequences to agent state changes

4.5.1. Creating behavior primitives

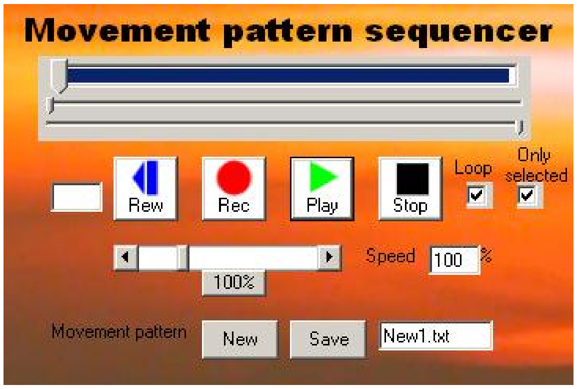
At the core of the animatronics control software is the **Manual Servo Control** (Figure 41), which allows the character designer to manipulate each DOF separately via sliders. In order to find the center, an additional Center button is provided per channel.



Servo 1: 91
Servo 2: 91
Servo 3: 91
Servo 4: 91
Servo 5: 91
Servo 6: 91

Figure 41: Manual servo control

The manipulation of DOFs is used in the **Movement Pattern Sequencer** (Figure 42), where behavior primitives are created and modified. Standard mode for recording primitives is a loop of 8 seconds, with a sample rate of 40Hz. The character designer modifies the position of the servos via the sliders in real-time. All changes are recorded automatically 'on the fly', and played back during the next loop. If a change is not satisfying, the designer can easily undo it by 'over-writing' the change during the next loop. This recording metaphor is similar to the 'audio dubbing' method used in movie making, where the actor watches a short scene in a loop, and can keep recording and adjusting the dubs until satisfaction.



Loop ☑
Only selected ☑
100%
Speed 100 %
Movement pattern New1.txt

Figure 42: Movement pattern sequencer

Creating primitives in a simultaneous playback/recording loop has proven to be a fast and efficient method. The same paradigm is used widely in musical sequencing software. This kind of behavior creation via direct manipulation may also be related to the 'programming by example' paradigm: in our context, the user teaches the system the desired behavior (by manipulating the sliders), and in a tight loop gets feedback of the system's performance by seeing both the sliders repeat what the character designer just did, as well as the animatronics following the slider movements.

In addition to direct manipulation via sliders, the character designer has access to each individual data point via conventional text editing, which guarantees maximum control over the behavior design process.

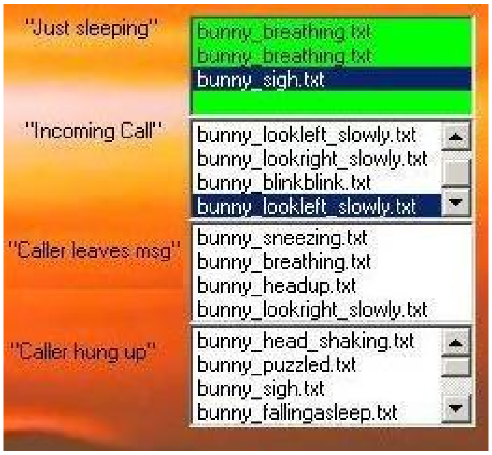
A movement primitive can be fine-tuned by reducing (or increasing) the speed of the loop recording and playback, which allows for finer control during the recording process. Furthermore, a primitive might start out as a 8-second loop, but can easily be pruned to a sub-section of the whole sequence by modifying the start and end points of the pattern; this pruning is done in a non-destructive way, and can be modified at any time. Once a primitive is built and modified to the designer's satisfaction, it can be stored in the **Movement Pattern Library**, and recalled at any time.

4.5.2. Composing complex behaviors

On the next level, the behavior primitives that are stored in the library can be composed into behavior sequences. Essentially, a behavior sequence consists of linearly arranged primitives; the software allows rapid creation of such sequences by simply dragging and dropping primitives into a list of other behaviors. Such a composited behavior sequence is stored, and can be played back in three modes:

- Play back whole sequence once, and then stop
- Play back all, and then repeat the last primitive
- Repeat whole sequence until the next behavior command is issued

4.5.3. Mapping behaviors to agent states



"Just sleeping"
bunny_breathing.txt
bunny_breathing.txt
bunny_sigh.txt

"Incoming Call"
bunny_lookleft_slowly.txt
bunny_lookright_slowly.txt
bunny_blinkblink.txt
bunny_lookleft_slowly.txt

"Caller leaves msg"
bunny_sneezing.txt
bunny_breathing.txt
bunny_headup.txt
bunny_lookright_slowly.txt

"Caller hung up"
bunny_head_shaking.txt
bunny_puzzled.txt
bunny_sigh.txt
bunny_fallingasleep.txt

Figure 43: Mapping messages to sequences

Each state change of the conversational agent may trigger behaviors of the animatronics. The cues are high-level descriptions of the agent state, such as "call received", or "caller finished recording a voice instant message," and are mapped to composite behaviors designed by the character designer. For each different animatronic device, the high level cues from the conversational agent are implemented according to its affordances (degrees of freedom, etc). This architecture allows an abstraction of the high level states of the conversation from the implementation of the respective behaviors in the animatronics. Therefore, animatronics with different affordances can get plugged into the same conversational system without the need to adjust the decision tree. This means that a user can choose which embodiment fits his/her mood, social setting, etc., without having to modify the conversational agent state machine, and lends new meaning to the phrase interface "skins."

The animatronics' behaviors are generated in real-time, depending on the agent-caller interaction. Therefore, factors such as the length of a voice instant message influence the animatronics behavior dynamically.

To create such dynamic behaviors, the conversational agent sends short messages to the animatronics server requesting certain behavior sequences when state changes occur. In addition, the agent can also specify the mode ('play sequence once', 'repeat all', 'repeat last primitive'), and the overall speed for the behavior. If a sequence is requested in 'repeat all' or 'repeat last primitive' mode, the animatronics repeats the behaviors until it receives a new command so the animatronics does not 'freeze' at the end of a sequence.

4.5.4. Interaction example

The example below shows the relationship between state transitions, the intended animatronics' behavior, and the low-level physical gestures (shown in parentheses). Although the example is fictitious, the current system works as described.

Joe is in a meeting. His animatronics, a palm-sized bunny with soft furry skin, is sleeping quietly. It is completely curled up, head tucked between its legs, eyes closed firmly and covered by its floppy ears (a). Every now and then it sighs (moves head twice up and down, 10% of actuator travel) in order to let its owner know that every-thing is ok, it's just asleep. A call comes in, and the bunny twitches slightly in its sleep, as if it had a dream (two sharp head movements, left-right-left-right to 20%, eyes opening 10% then closing again), but is still asleep (b). The Intermediary then recognizes the caller from caller ID: it's Joe's friend Clara. The bunny sighs, and slowly wakes up (slow head movement up and 30% to the left; at the same time, its eyes start to open slowly to 50%, close again, open twice for 20%; the head shakes slightly left-right-left, then the eyes open, a bit faster now, to 70%, (c).

The agent asks Clara if she wants to leave a voice mail or voice instant message. Clara leaves a voice instant message. During that time, the bunny sits still, looks up as if it would listen to something only it can hear, slowly turning its head from left to right, blinking once in a while (d). As soon as she is done leaving the message, the bunny gets excited and looks around pro-actively (rapid full movements of the head from one side to another). Joe notices it, and turns his attention towards it (e). The bunny whispers in his ear and tells him who is on the phone, then plays back the short message it took from Clara (f). The animatronics is now fully awake and attentive (eyes completely open, head straight) (g). Joe touches the bunny's right ear (which triggers the recording mode) to leave a reply. The bunny sits still, listening (head tilted slightly upwards, blinking fast and of-ten) (h). As soon as Joe is *done, it confirms by nodding (medium fast head movement down and then back to middle, followed by single blink). When the message has been delivered to Clara, the bunny looks back at Joe and winks at him, to confirm the delivery (head straight, one eye blinks twice). Then it stretches (head slowly upwards to 100%, then medium fast back to middle), and gets sleepy again (eyes close to 50%, and slowly closing and opening again, twice; at the same time, the head goes slowly down to its belly, halting 2 times in the movement), eventually assuming the same curled up posture it had before the call.*

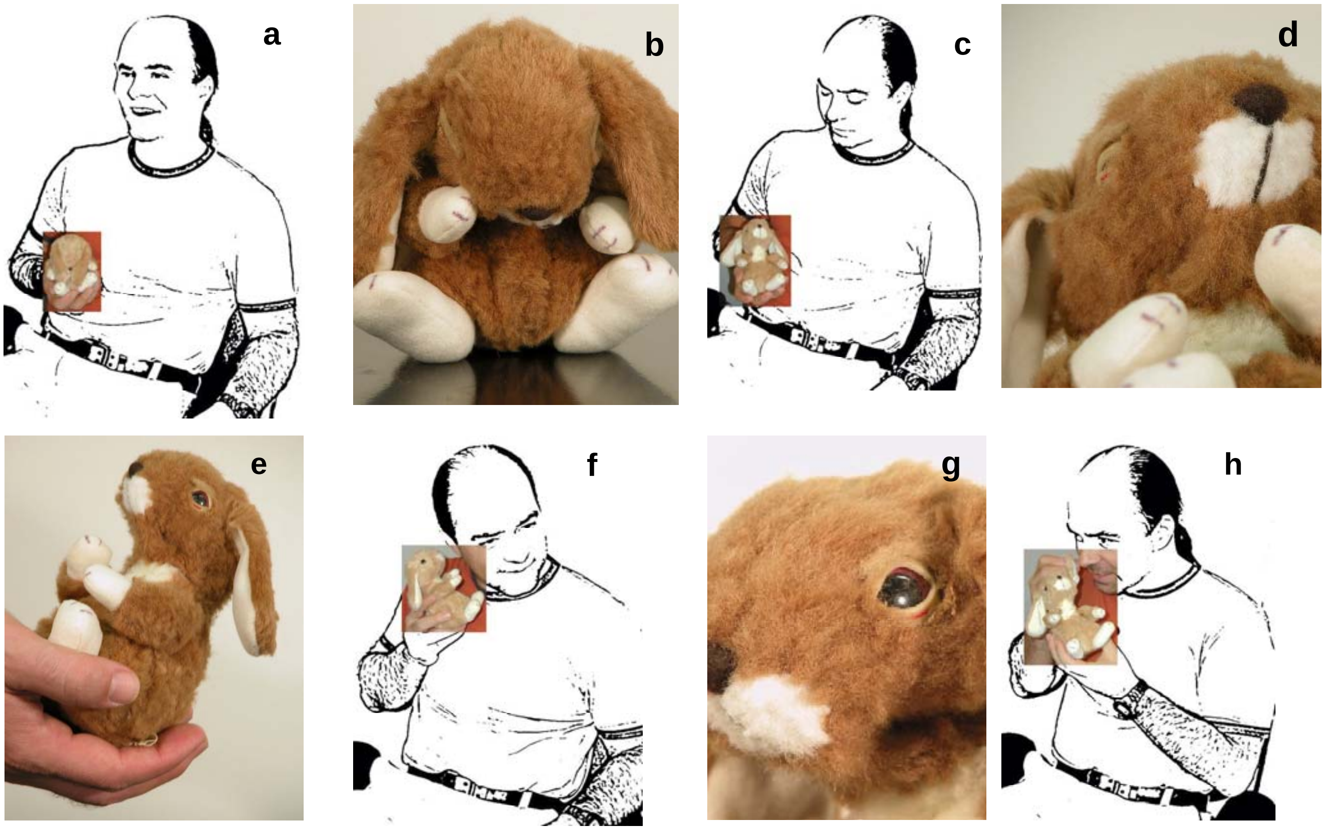


Figure 44: top row: bunny sleeping, waking up, listening to caller
bottom row: trying to get attention with gaze, whispering to user, being attentive, listening to user

4.6. Conversation Finder

The purpose of the Conversation Finger subsystem is to provide the Intermediary with information about the conversational status of the user. This is achieved by utilizing a decentralized network of autonomous body-worn sensor nodes. These nodes detect conversational groupings in real time, and offer the Intermediary information about how many people participate in the user's conversation, as well as if the user is mainly talking or listening.

Each user owns his or her Conversation Finder node, worn close to the neck. It functions as binary speech detector and communicates asynchronously with other nodes on a single radio channel. Each node sends out frequent heartbeat messages over RF, as well as a message when the user is talking, and receives messages from the nodes that are close by. The nodes independently come to a conclusion about who is in the user's current conversation by looking at alignment and non-alignment of the speaking parties. At any time, the Intermediary can query the user's node wirelessly for this continuously updated list of people.

Each node consists of two double-sided PCB boards with two PIC 16LF877 microcontrollers, microphone capsule, Radiometrix© Bim2 transceiver (in the 433MHz spectrum), microphone preamplifier, and a 140mAh lithium polymer battery. The overall size is 40x35x20mm.

4.6.1. Conversational groupings

In order to detect conversational groupings, the Conversation Finder nodes assume that if two people are in a conversation with each other, their speaking does not overlap for a significant amount of time (Basu, 2002) [9]. A “significant amount of time” may be a culturally biased parameter, but an overlap of 3 seconds has proven to be a useful value in informal tests.

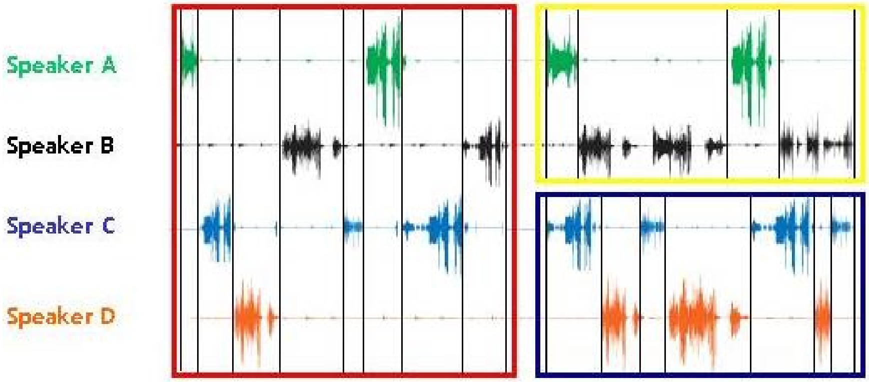


Figure 45: Alignment of speech: on the left side (red area), all four speakers’ speech signal is aligned, so they are probably in a single conversation. On right side, speaker A and B are aligned, and C and D, which probably means that these are two separate conversations.

4.6.2. Simulations

In order to test the messaging protocol, some simulations were done prior to implementing the system.

A software simulation of the wireless sensor network was created to test protocols and algorithms. Each node is represented by a single computational process. Since there is not real speech involved, a

```
12345678
x   x
 x  x
 x  x
x   x
x x x
  x
x      x
x
x x x
 xx    x
 xx x
x x
x  x  x
xx x  x
 x x  x
 xxx  x
 xx
 xx    x
x x    x
x  x
x  x x
 x x x
```

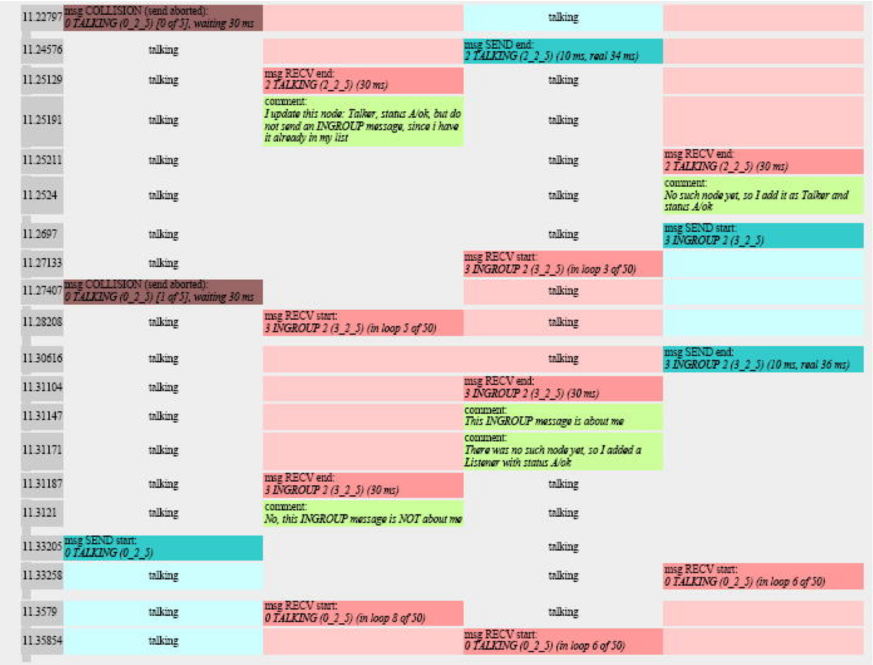
Figure 46: Conversation script for 8 nodes, used for simulation

'conversational script' is created that each nodes loads upon startup. Figure 46 shows an example for a conversation with eight nodes. The script defines when each node is about to 'speak.' Each line of the script corresponds to 1000ms (but can be adjusted for time lapse or high speed simulation), and an "X" marks the time windows each node is speaking.

A pacemaker process sends messages to each node when to advance to the next line in the script.

Each node then sends out messages (according to the script) and listens for incoming messages. During the simulation run, each node generates an extensive log file with time-stamped events: the messages it has sent, the messages it has received, its current status (the "conversation list"), comments, etc. The timestamps have millisecond resolution to show conflict, collisions, etc.

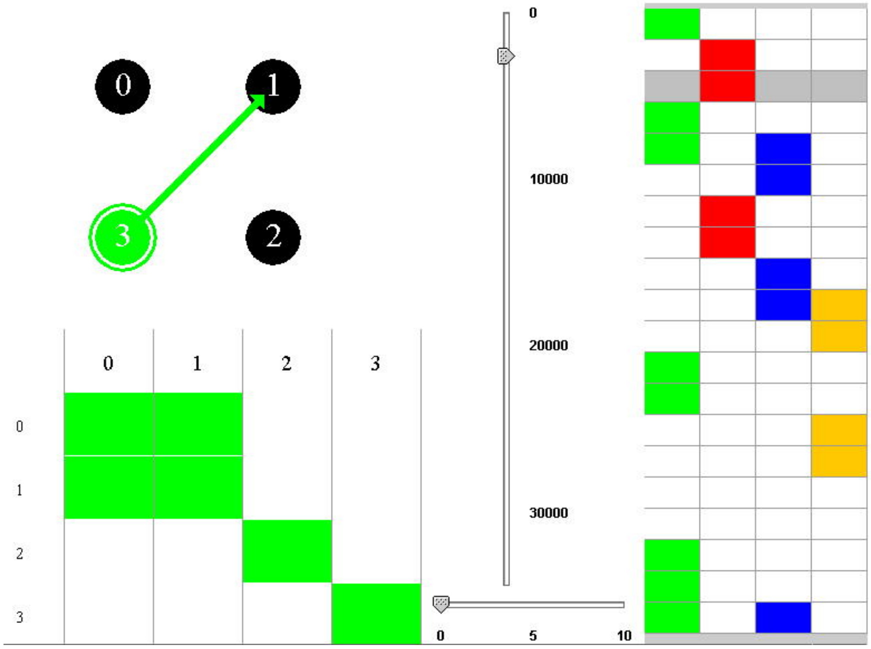
Figure 47 shows a single page out of 115, which was generated by a 34-second script of a 4-node conversation. Each column represents the log file of one node. The messages are color coded, and timestamps include the beginning of the message as well as the end. Since communication between the nodes was done by writing to a shared file space, collisions of message can be detected very clearly.



11.22797	msg COLLISION (send aborted): *0 TALKING (0_2_5) [0 of 5], waiting 30 ms*		talking	
11.24576	talking		msg SEND end: *2 TALKING (2_2_5) (10 ms, real 34 ms)*	
11.25129	talking	msg RECV end: *2 TALKING (2_2_5) (30 ms)*	talking	
11.25191	talking	comment: *I update this node: Talker, status A/ok, but do not send an INGROUP message, since i have it already in my list*	talking	
11.25211	talking		talking	msg RECV end: *2 TALKING (2_2_5) (30 ms)*
11.2524	talking		talking	comment: *No such node yet, so I add it as Talker and status A/ok*
11.2697	talking		talking	msg SEND start: *3 INGROUP 2 (3_2_5)*
11.27133	talking		msg RECV start: *3 INGROUP 2 (3_2_5) (in loop 3 of 50)*	
11.27407	msg COLLISION (send aborted): *0 TALKING (0_2_5) [1 of 5], waiting 30 ms*		talking	
11.28208	talking	msg RECV start: *3 INGROUP 2 (3_2_5) (in loop 5 of 50)*	talking	
11.30616	talking		talking	msg SEND end: *3 INGROUP 2 (3_2_5) (10 ms, real 36 ms)*
11.31104	talking		msg RECV end: *3 INGROUP 2 (3_2_5) (30 ms)*	
11.31147	talking		comment: *This INGROUP message is about me*	
11.31171	talking		comment: *There was no such node yet, so I added a Listener with status A/ok*	
11.31187	talking	msg RECV end: *3 INGROUP 2 (3_2_5) (30 ms)*	talking	
11.3121	talking	comment: *No, this INGROUP message is NOT about me*	talking	
11.33205	msg SEND start: *0 TALKING (0_2_5)*		talking	
11.33258	talking		talking	msg RECV start: *0 TALKING (0_2_5) (in loop 6 of 50)*
11.3579	talking	msg RECV start: *0 TALKING (0_2_5) (in loop 8 of 50)*	talking	
11.35854	talking		msg RECV start: *0 TALKING (0_2_5) (in loop 6 of 50)*	

Figure 47: Small excerpt of a conversation finder simulation log file

Although the log files of all involved nodes taken together would explain the behavior of the system, it is very difficult to understand what is going on. Therefore, an animated visualization of these log files was created to trace failures of the system, and fine-tune the protocol and algorithms. Figure 48 shows a screenshot of the interface: on top left, there are four nodes shown. At that point in the simulation Node 3 is sending a message to node 1. On the right side, the conversational script of the four nodes it depicted. In between is a slider with a timeline function that allows the user to jump to any point in time of the simulation. On the lower left side of the interface is the conversation matrix, depicting each node's memory content.



	0	1	2	3
0	🟩	🟩		
1	🟩	🟩		
2			🟩	
3				🟩

Figure 48: User interface of the animated visualization of the conversation finder simulation results (by Quinn Mahoney)

4.6.3. Messaging protocol

After many trials, a messaging protocol was developed that was simple yet efficient. Each message consists of one byte (repeated for error checking purposes). The first nibble is the message ID; the second nibble is the node ID.

Each node sends out a HEARTBEAT message every 3000ms. When the wearer of a node is talking, the node sends out TALK messages continuously, 6 every 200ms.

A 4 bits message space and 4 bit ID space allows for 16 different kinds of commands, as well as 16 different node IDs. The complete messaging protocol is shown in section 4.8.3. The flow chart of the firmware, describing each node's behavior when messages arrive, will be described later in this section.

4.6.4. Circuit design, breadboards

A Conversation Finder node consists of two main elements: an audio part with a microphone, amplifier and microcontroller to analyze the microphone signal, and a transceiver part with the radio module and yet another microcontroller.

The audio part amplifies the microphone signal, then the controller digitizes it with 10 bits, integrating it over time and providing the transceiver part with a single bit of information about if the user is talking or not.

Figure 49 shows the schematic of the audio system, Figure 50 the schematic of the transceiver system.

Both parts of the node are based on PIC 16LF877 microcontrollers. The processors used are able to run with voltages as low a 3V, a prerequisite for using 3.7V lithium polymer cells. As a consequence, the controllers can be clocked with only 4MHz, which in turn limits the maximum serial speed to 19.2kbps.

The transceivers in the nodes are Radiometrix© BiM2, which operate in the free 433MHz spectrum, and have an output of 10dBm (10mW) nominal that gives them a range of about 20 meters indoors. Used were special low voltage versions that have no problem with a single lithium polymer cell's voltage. On the breadboards, there are 16cm long wire antennas, a quarter of the wavelength of 433MHz. In the PCB version, the antenna is integrated as a trace.

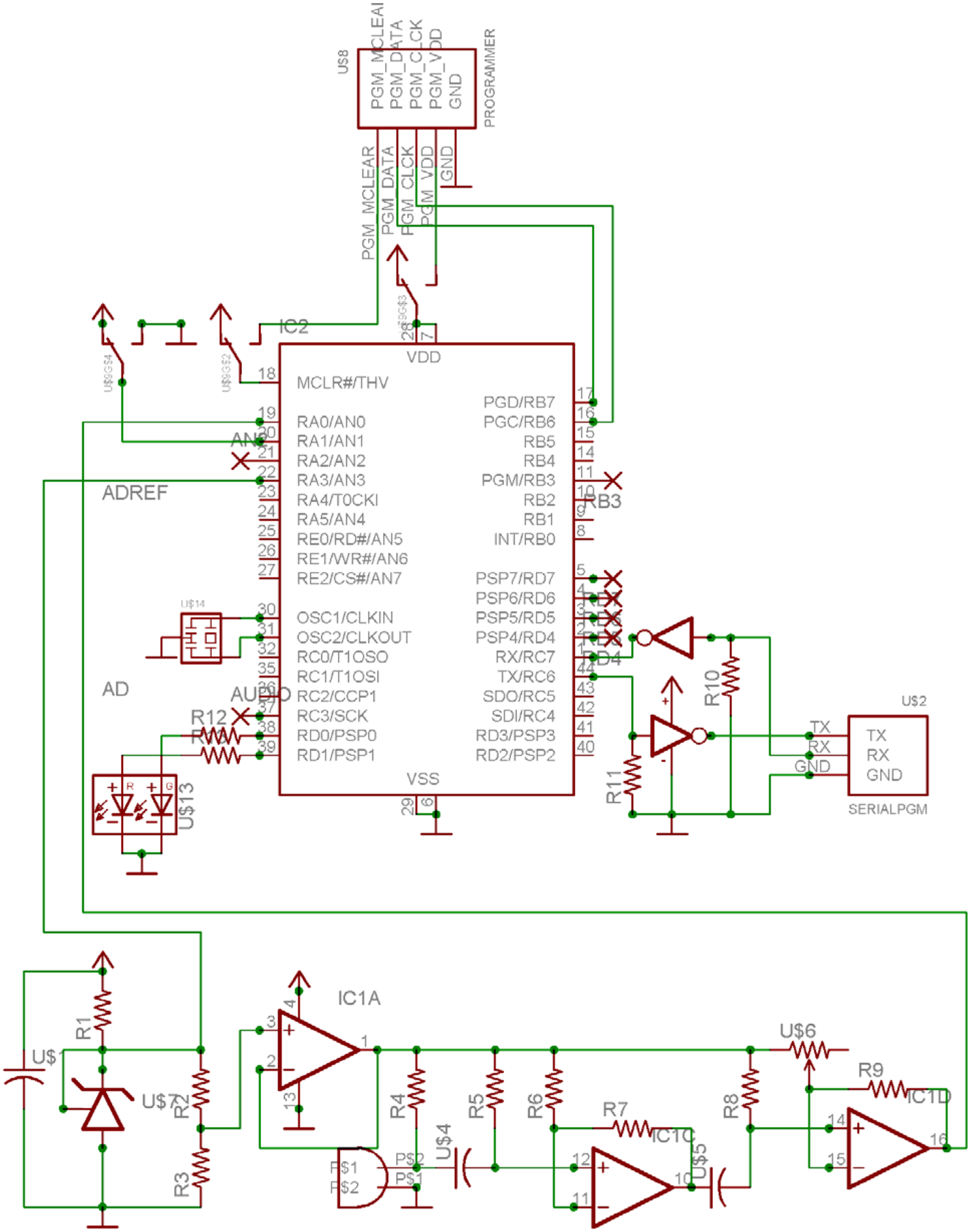


Figure 49: Schematic of audio board

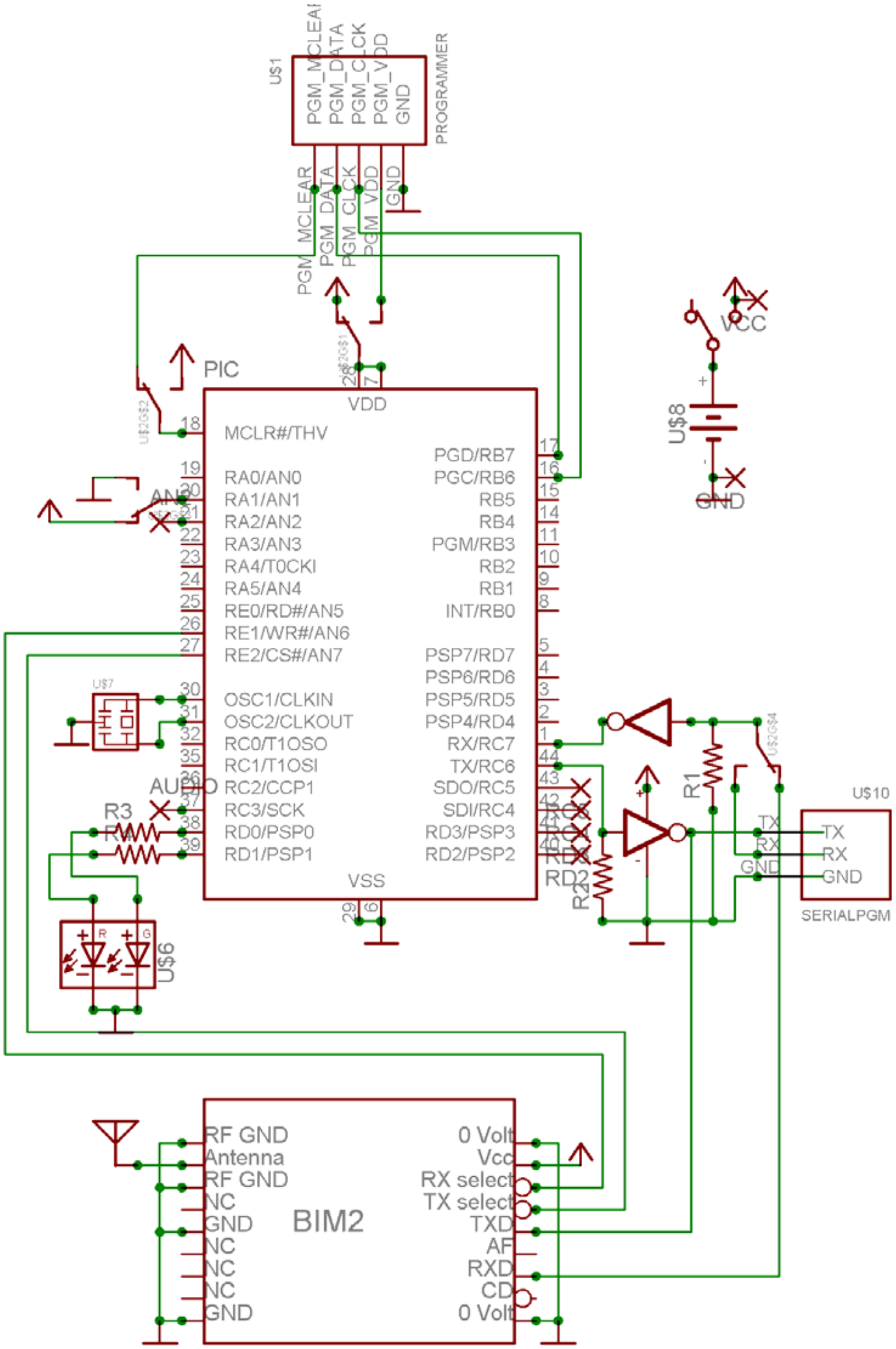


Figure 50: Schematic of transceiver board

All components were set up on solderless breadboards (Figure 51). These two boards were used to fine-tune all component values and test the initial software for both microcontrollers.

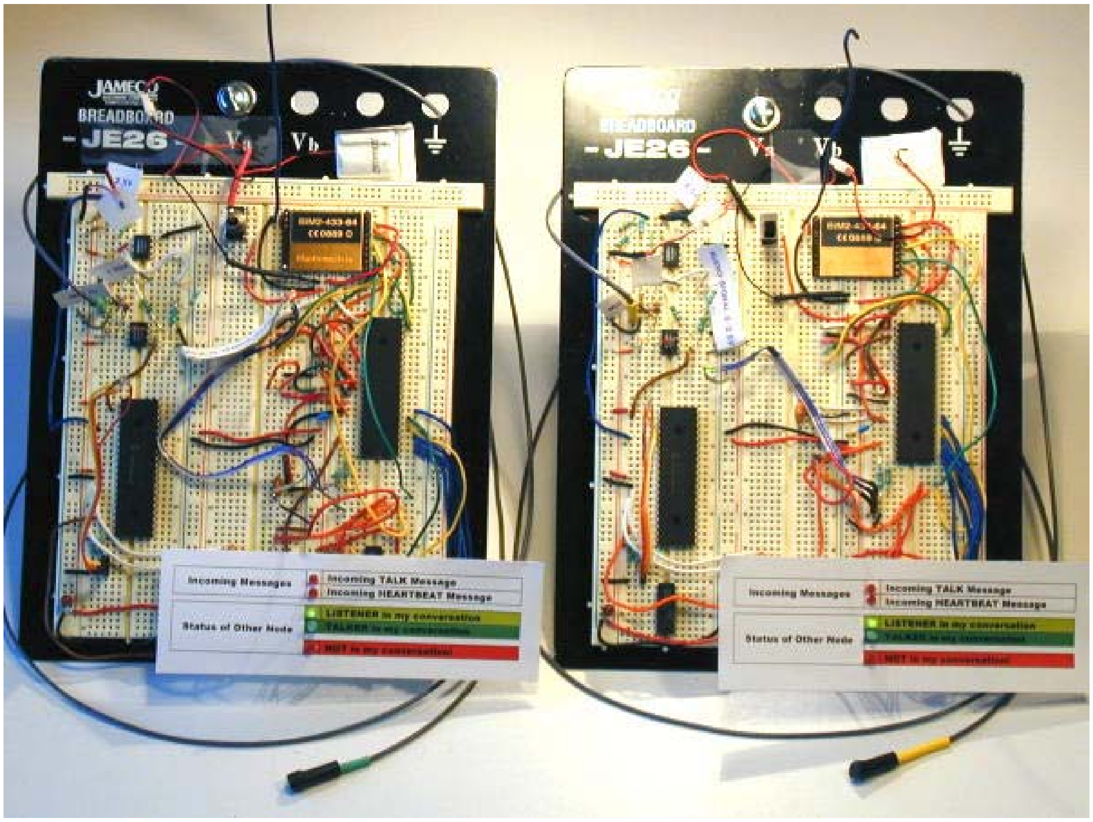


Figure 51: Breadboards of initial Conversation Finder nodes

The breadboards were equipped with additional status LEDs that show what each node 'thought' of the other one: if it was visible (HEARTBEAT received), if it was a listener or talker, or if the other node is thought to be part of another conversational grouping.

4.6.5. Software

There are two microcontrollers per node that have to be programmed. Initial programming is done with a Picstart Plus development programmer, which takes about 5 minutes. During this step, a boot loader routine is installed. Most (but not all) subsequent programming is done via inline serial programming, which takes only a few seconds.

All software is written in C (with a few assembly lines include), and then compiled with a CCS compiler.

Audio node code
The audio microcontroller's code is identical for all nodes. In a loop, it adds up one thousand 10bit samples (which takes 183ms, resulting in a sampling rate of 5.45kHz). It then calculates the average value, and raises the talk line in case it is above a certain threshold. In addition to this software threshold, each audio board also contains a potentiometer to adjust the analog amplification level of the microphone preamp.

Transceiver node code

Each transceiver node contains identical code as well, except for its node ID. The code is more complex since it manages the transmission and receptions of RF messages, and continuously updates its internal data structure that describes the status of the other nodes as well as the user's conversational status.

The node's main program consists of a loop that lasts about 200ms, and contains the following steps:

- Listen for incoming messages for about 200ms
- If the user is talking, send out a TALK message
- Update internal data structure
- Keep track of the user's "talk-to-listen" ratios
- Send out a HEARTBEAT message (every 3000ms)

The logic of the transceiver node in terms of its internal data structure is as follows: Each node listens for incoming radio messages from nearby nodes. Upon receiving a 'heartbeat,' the other node is classified as *Listener*. Detecting a 'talk' message will upgrade its status to a *Talker*. Each node continuously determines if the detected nodes might be part of its owner's conversation or not. If the node's microphone determines that its user is talking, and simultaneously receives 'talk' messages from another node for more than a three-second window, it excludes the other node for a 30-second period by tagging it as *Excluded*. If a node classified as a *Talker* stops sending 'talk' messages, it will get re-classified to a *Listener* after a period of time. Similarly, if a node fails to send out 'heartbeat' messages, it will get tagged as *Absent* by the other nodes. This continuous process of classifying all other nodes is done in each sensor node independently, and during informal tests with a set of six prototype nodes, this logic demonstrated to be a reliable and fault tolerant source of conversational status information.

The transceiver node also continuously calculates how much the user is talking, versus being quiet or listening. It does so for three different time periods (rolling windows): the last 3.2 seconds, the last 51.2 seconds, and the last 819.2 seconds. The Intermediary can poll these values, providing it with important information about the user's conversational status.

Calculated are these "talk-to-listen" ratios from three hierarchical levels of circular audio buffering (Figure 52). Each buffer's overall result is piped into the next higher buffer's basic slot:

- *First-level buffer*: 16 slots (bits), each representing 0.2 seconds. If there was talk activity during the last 200ms segment, a bit of the first-level buffer is set to high. This first-level buffer covers the last 3.2 seconds.
- *Second-level buffer*: 16 slots, each representing 3.2 seconds. If the last first-level buffer (3.2 seconds) contained *any* talk activity (any of the 16 bits set to high), a bit of the second-level buffer is set to high. This second-level buffer covers the last 51.2 seconds.
- *Third-level buffer*: 16 slots, each representing 51.2 seconds. If the last second-level buffer (51.2 seconds) contains more than 50% talk activity (more than 8 of the 16 bits set to high), a bit of the third-

level buffer is set to high. This third-level buffer covers the last 13 minutes 39.2 seconds.

Each of the three buffers describes its talk time percentage with a resolution of 4 bits (16 values). An example for how the Intermediary polls this information can be found in section 4.8.2.

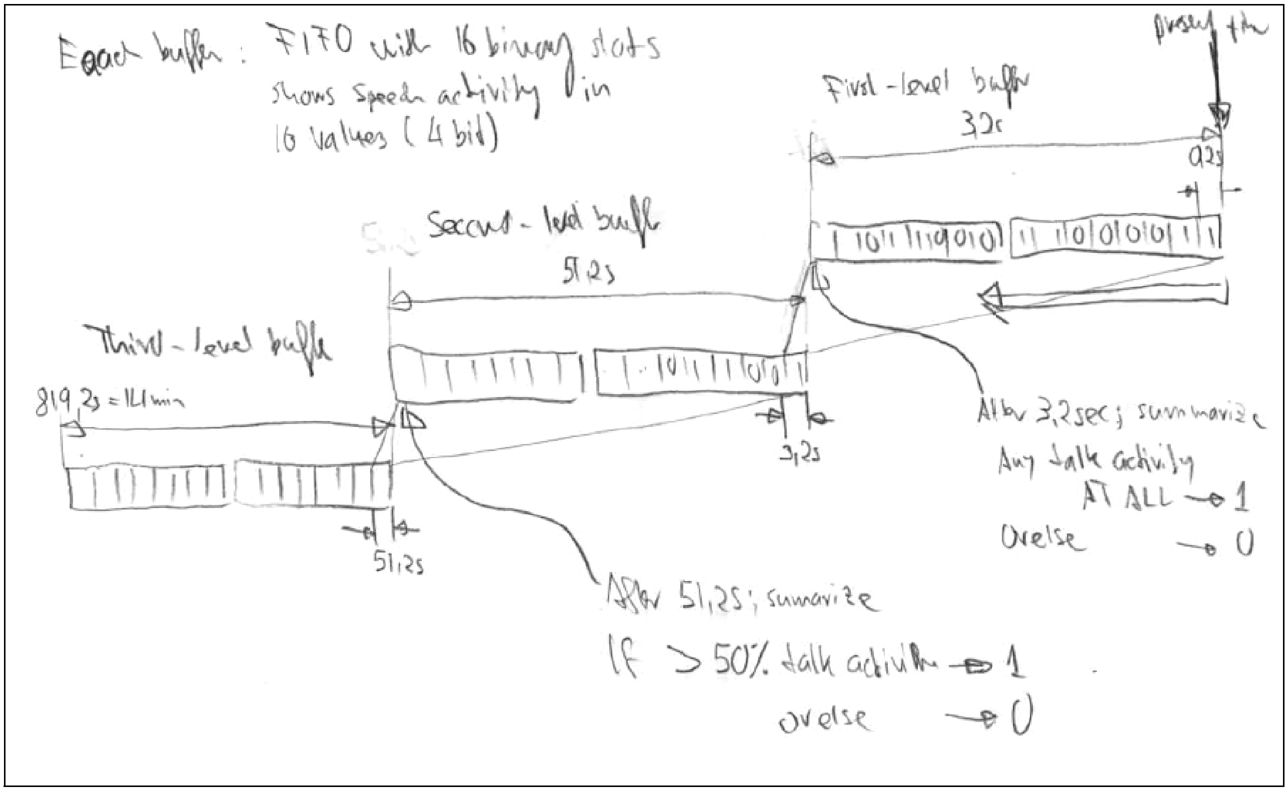


Figure 52: Sketch of algorithm to calculate the user's talk-to-listen ratios

The nodes do not listen only for messages from other Conversation Finder nodes, but also for messages from the sensor network hub, as well as for special messages that are used for debugging the sensor network. In addition to TALK and HEARTBEAT messages, the nodes also send other information, such as commands to the Finger Ring nodes to vibrate, and other types of information.

The complete set of messages is described in section 4.8.

4.6.6. PCBs

After the two initial breadboard nodes were working properly, surface mount versions were developed.

As mentioned earlier, each conversation finder node consists of two double-sided PCB boards with two PIC 16LF877 microcontrollers, microphone capsule, Radiometrix© Bim2 transceiver (in the 433MHz spectrum), microphone preamplifier, and a 140mAh lithium polymer battery. The overall size is 40x35x20mm.

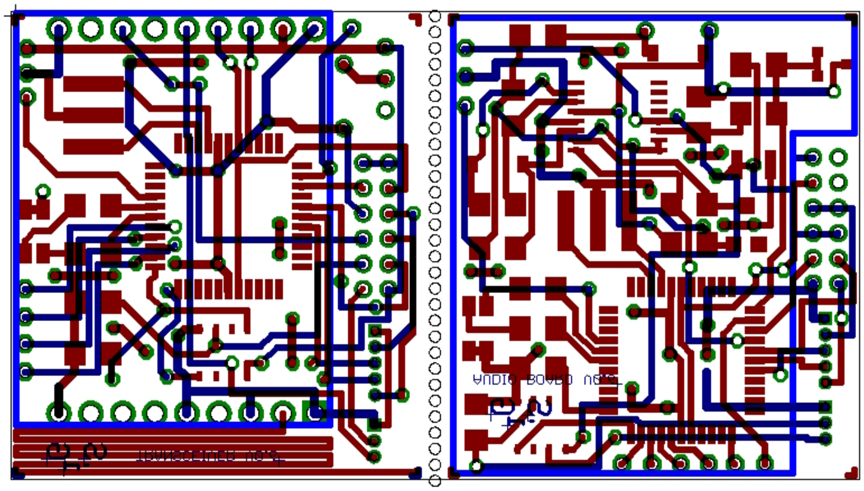


Figure 53: Conversation finder boards in EagleCAD

The boards were manufactured without silk screen, and came back like shown in Figure 54.

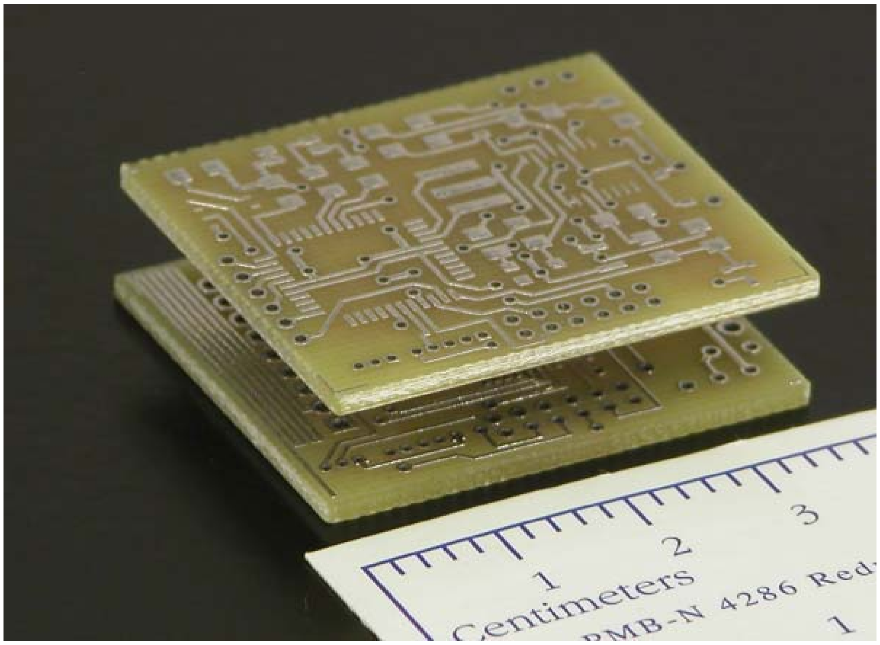


Figure 54: Raw conversation finder PCBs

Then all the surface mount components were soldered onto the boards manually (Figure 55).

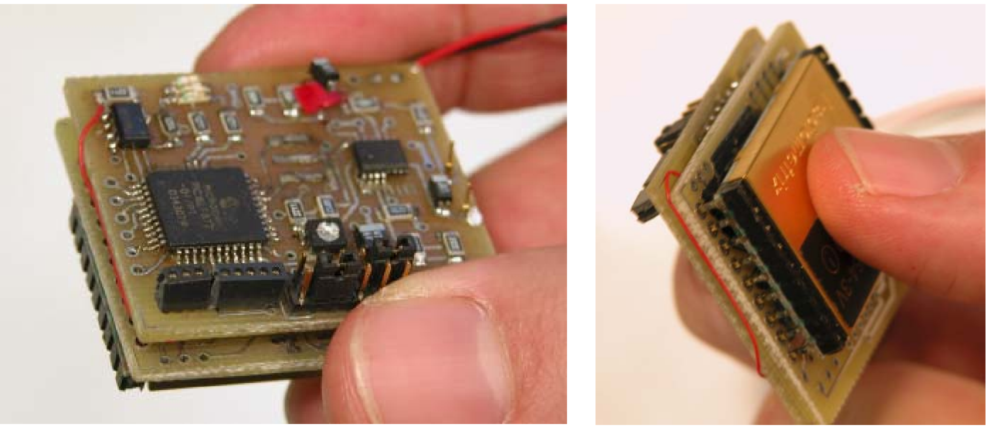


Figure 55: Finished Conversation Finder node

The two boards are connected via a three-pin header (voltage, ground, and talk signal). Microphone capsules with wires of different lengths were used to test the best position of the node on the body of the user.

Informal tests showed that the ideal position is inside the user's shirt, right in the middle of the neck opening (under the chin). A short microphone cable is used to point the capsule towards the neck.

4.6.7. Packaging

In order to wear or attach them, the nodes were fit into a pocket made of stretchy cloth. On the back of the node is either a safety needle (Figure 56), or the whole node is suspended around the user's neck with a necklace.

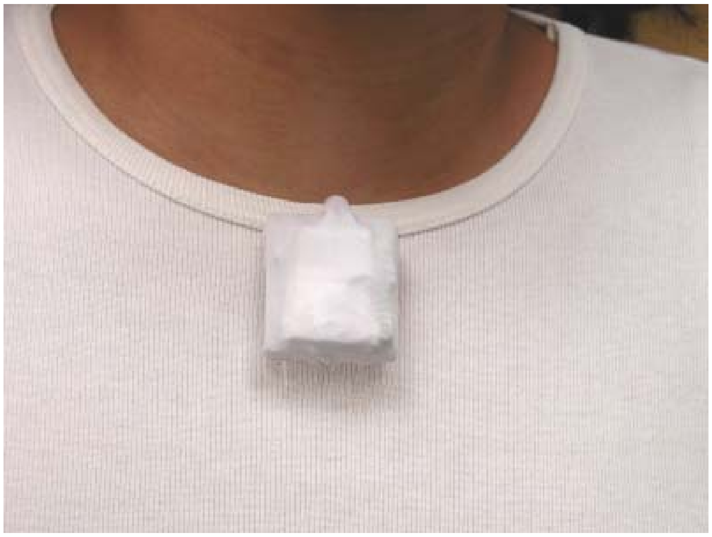


Figure 56: Conversation finder node attached to shirt

4.7. Finger Ring

The actuated ring consists of a tiny vibration motor (pager motor with excenter), a 20mAh lithium polymer battery, a micro switch, Radiometrix Bim2 transceiver (operating in the 433MHz spectrum), and a PIC 16F877 microcontroller.

The Finger Ring's transceiver receives messages from its user's Conversation Finder node indicating that it has to alert the ring wearer, upon which it vibrates slightly. If the user touches the micro switch located under the ring, the transceiver broadcasts a veto message to the Intermediary. For user testing, wired versions of the Finger Ring were built.

4.7.1. Messaging protocol

Although the Finger Ring nodes are part of the Intermediary's sensor network and use the same transceivers as the Conversation Finder nodes, each Finger Ring node only looks for one message type: a message from its Conversation Finder node asking it to vibrate.

This message is called CONTRACT (for legacy reasons), and contains a target ID. If the node receives such a message, it compares the target ID with its own ID. If there is a match, the microcontroller turns on the vibration motor for 1000ms. During trials, this value has been proven to be subtle enough not to interrupt, but still perceivable by the wearer.

After the reception of a valid CONTRACT message, a 10-second window opens. If the user decides to veto to the upcoming interruption, she has ten seconds to press the micro switch attached to the under side of the ring. If she decides to veto, the ring broadcasts a VETO message. This message is anonymous, but contains as a payload the ID of the interrupting agent. This allows for several polling processes at the same time. Therefore, the requesting agent can see if an incoming VETO message is meant for it, but does not know its origin.

If a user presses the micro switch on the ring outside this 10-second window (before or after), a different message (VETO_OWN) is sent out which is addressed specifically to the Finger Ring's own Intermediary. This is done so that the user can use the finger ring for other purposes, like to influence the animatronics, or to pick up an incoming call. To the Intermediary, it is perceived as a button press, similar to the switches in the extremities of the animatronics.

A complete list of all commands can be found in section 4.8.

Like in the Conversation Finder nodes, the code is written in C, compiled with CCS, and programmed onto the PIC with a Picstart Plus development programmer.

The code runs as a loop with the following elements:

- Listen for incoming messages for 200ms, and keep track of the user's button presses
- Send out a veto message if the user has pressed the button
- Send out a HEARTBEAT message (every 3000ms)

4.7.2. Circuit design, breadboard

The circuit design shares many features with a Conversation Finder transceiver node. At the core of a finger ring node is a PIC 16LF877 microcontroller and a BiM2 transceiver, both low-voltage versions (3V) that can be powered with a single 3.7V lithium polymer cell.

Additional components are an inverter (for the serial signals), an oscillator, a status LED, a transistor to switch the motor, the motor with excenter, and a battery.

The basic components are set up on a solderless breadboard (Figure 57).

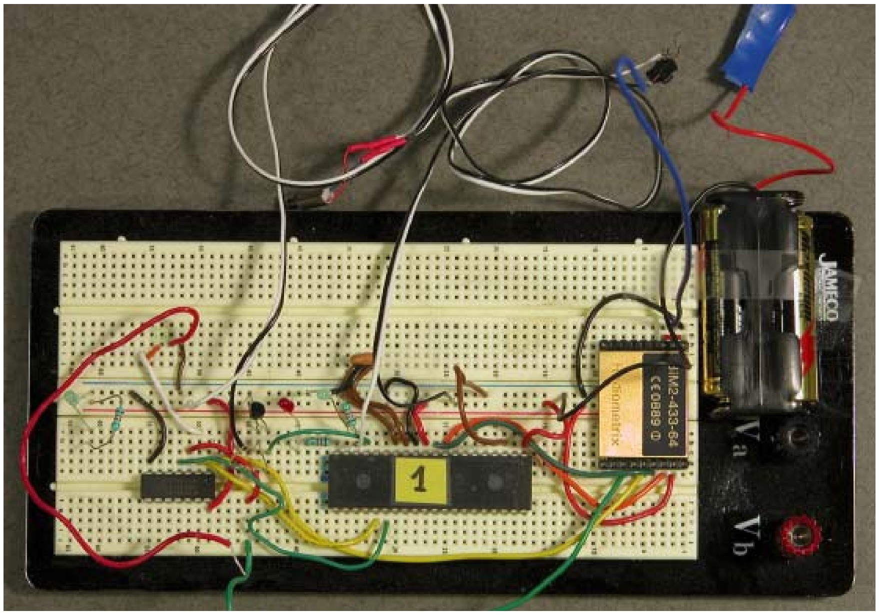


Figure 57: Breadboard setup of a finger ring node

4.7.3. Wireless prototype

The footprint of the BiM2 transceiver determines the overall size of the finger ring node. All electronics elements fit underneath the transceiver.

There is neither a PCB board (like with the Conversation Finder nodes) nor a perforation board (like with the animatronics controller board). Instead, all components are soldered directly together with wires: especially the microcontroller's connections were a challenge, with each pin being less than half a millimeter wide, and even less than that apart from the next pin (Figure 58).

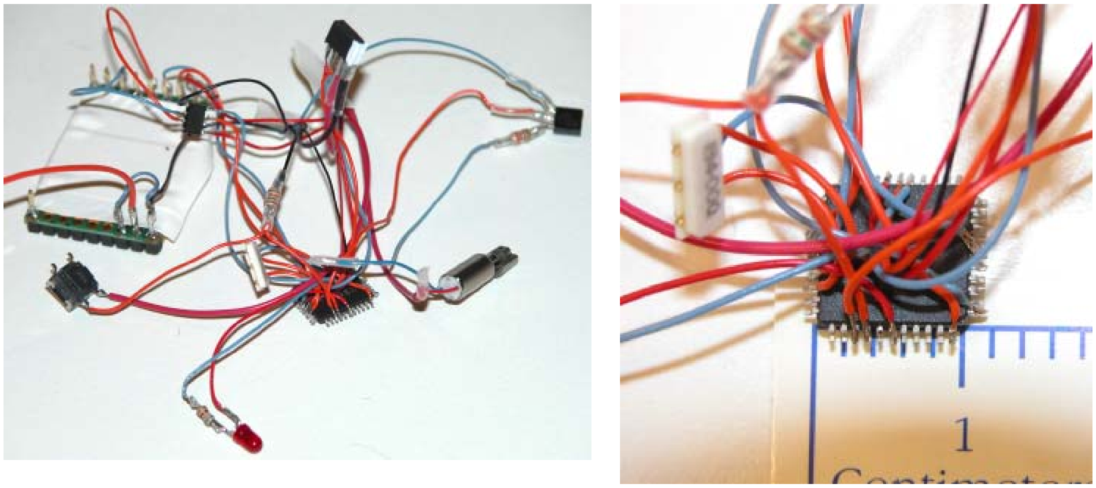


Figure 58: Finger ring node components (left), microcontroller with direct wires (right)

After the successful completion of this procedure, all the components of the “spider web” are carefully packaged into the lower section of the BiM2 transceiver (Figure 59). A metal ring is attached to the backside of the transceiver, and the micro switch attached to the side of the ring. Although the ring appears larger than average jewelry, it is still well within the range of fashionable finger rings available these days.

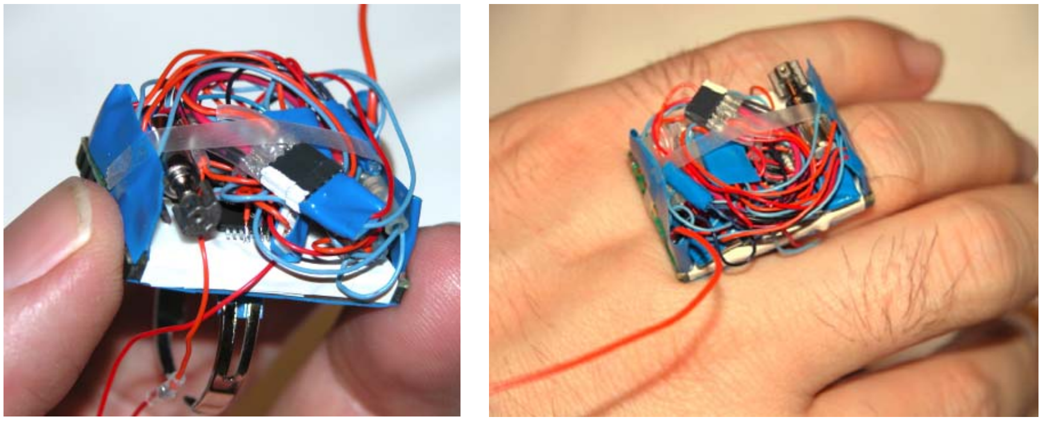


Figure 59: Finger Ring node

4.7.4. Wired prototype

Prior to the wireless finger rings, wired rings were built with similar vibration actuation as the wireless ones (Figure 60). These rings consist of a ring with an attached flat excenter motor (disk motor), diameter 13mm, and a micro switch. There are no other electric or electronic components involved.

A thin multiconductor cable connects each ring to a central ring station where vibration is triggered manually, and button presses of the ring wearers are indicated with LEDs of different colors.

These wired rings are used for an experimental study, described in section 5.2.

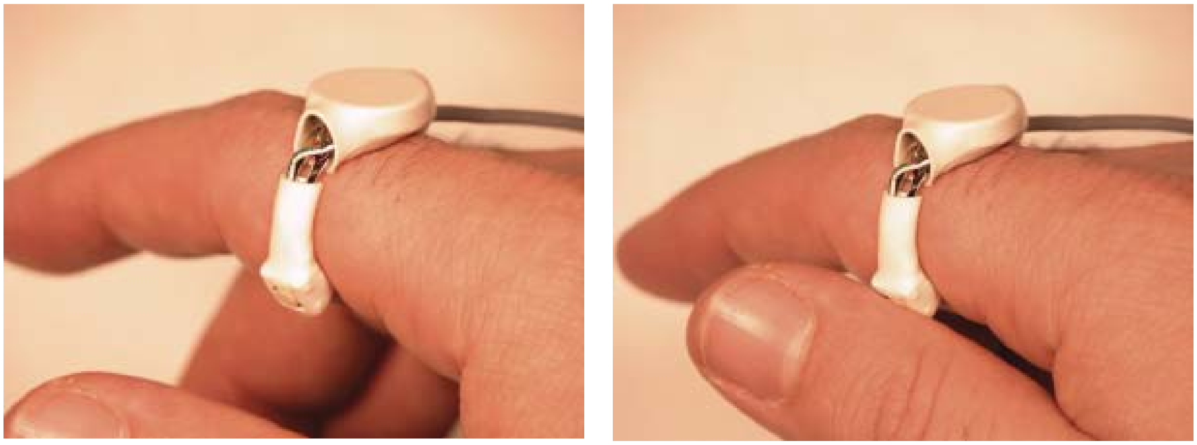


Figure 60: Wired finger ring for user testing

4.7.5. Interaction Conversation Finder - Finger Ring nodes

In our wireless system, the network protocol hides the identity of the vetoing person from the phone that queries all participants for their input (Figure 61).

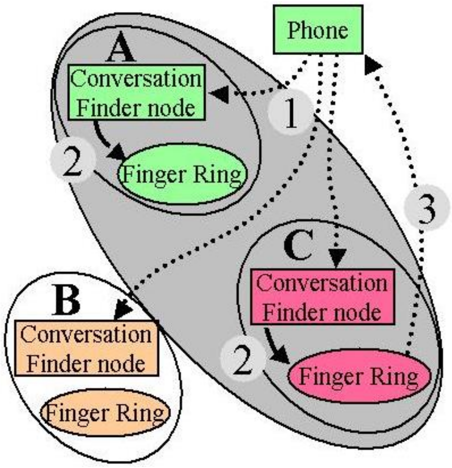


Figure 61: Interaction Conversation Finder – Finger Ring nodes

The querying phone of user A broadcasts a message to all Conversation Finder nodes in range (1). If they ‘think’ they are in a conversation with user A (in our example: user A and C, but not user B), they are asked to send a directed message to their respective finger rings, which will cause these rings to vibrate as a pre-alert (2). At that point, all participants who received pre-alerts (A and C) can veto the upcoming interruption by pressing the finger ring's micro switch. If a user presses the switch (such as user C), the ring will broadcast an anonymous veto message that will be picked up by the querying phone of user A (3). Note that A’s phone never knows who else is in a conversation, much less who vetoed (but can still count the number of vetoes received).

A comprehensive list of all commands can be found in section 4.8.

4.8. Sensor network hub

All nodes of the sensor network are perfectly able to function on their own, since they are conceived as an adhoc, decentralized network. They are built to interact mainly with each other.

However, the Intermediary software is running on a remote PC and needs to communicate with its sensor network somehow. For this purpose, a sensor network hub was build that connects to the serial port of a PC and can interact with the nodes of the sensor network.

4.8.1. Hardware

The hardware involved for the sensor network hub is a BiM2 transceiver connected to a desktop computer. It consists of a small PCB board (made by Vadim Gerasimov) that houses the transceiver, as well as an RS232 cable (serial) for communication, and a USB cable for power (Figure 62).

This transceiver is identical to the transceivers used for the Conversation Finder nodes as well as for the Finger Ring nodes.

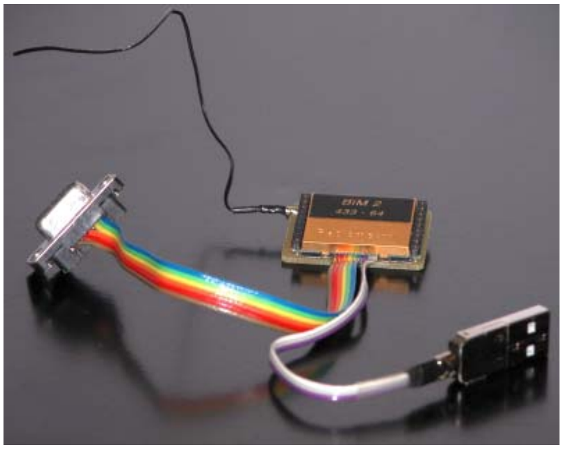
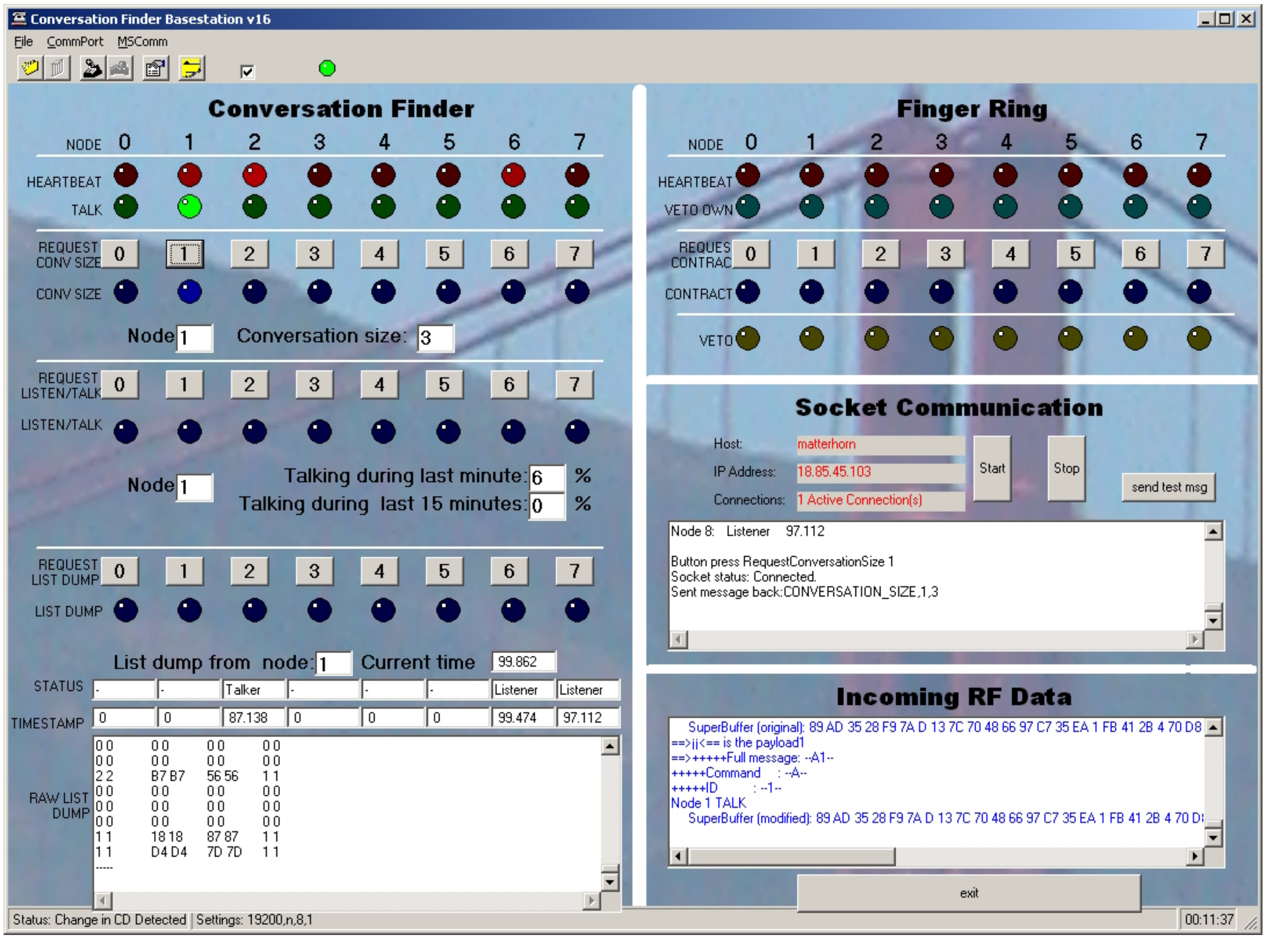


Figure 62: Sensor network hub transceiver

4.8.2. Software

The sensor network hub consists of a server that has a graphical user interface (Figure 63). Its main function, though, is to relay socket messages from the Intermediary to the sensor network nodes.



Node: 1 Conversation size: 3

Node: 1 Talking during last minute: 6 % Talking during last 15 minutes: 0 %

List dump from node: 1 Current time 99.862

STATUS	-	-	Talker	-	-	-	Listener	Listener
TIMESTAMP	0	0	87.138	0	0	0	99.474	97.112

```
0 0     0 0     0 0     0 0
0 0     0 0     0 0     0 0
2 2     B7 B7   56 56   1 1
0 0     0 0     0 0     0 0
0 0     0 0     0 0     0 0
0 0     0 0     0 0     0 0
1 1     18 18   87 87   1 1
1 1     D4 D4   7D 7D   1 1
-----
```

Status: Change in CD Detected | Settings: 19200,n,8,1

Host: matterhorn
IP Address: 18.85.45.103
Connections: 1 Active Connection(s)

```
Node 8:   Listener    97.112

Button press RequestConversationSize 1
Socket status: Connected.
Sent message back:CONVERSATION_SIZE,1,3
```

```
    SuperBuffer (original): 89 AD 35 28 F9 7A D 13 7C 70 48 66 97 C7 35 EA 1 FB 41 2B 4 70 D8
==>¡¡<== is the payload1
==>+++++Full message: --A1--
+++++Command     : --A--
+++++ID          : --1--
Node 1 TALK
    SuperBuffer (modified): 89 AD 35 28 F9 7A D 13 7C 70 48 66 97 C7 35 EA 1 FB 41 2B 4 70 D
```

00:11:37

Figure 63: Sensor network hub user interface

For debugging purposes, this software can monitor all activity going on in the RF channel, which includes the messages passed between the sensor nodes, both from the Conversation Finders and Finger Rings.

On the left side, all currently present Conversation Finder nodes can be monitored. Red LEDs light up when a HEARTBEAT message comes in, and green ones show when a TALK message arrives.

Right below these two rows, one can manually request the conversation size of each Conversation Finder node, as well as the current talk-to-listen ratio. On the lower left side, one can request a memory dump of each node, consisting of a list of nodes it sees, what these nodes are classified as (talker, listeners, excluded, etc), as well as timestamps.

On the right side, all present Finger Ring nodes can be monitored. The rings' heartbeat messages are displayed (red LED), as well as when a ring wearer presses her micro switch (VETO_OWN).

In the row below, one can manually broadcast a request to all present Conversation Finder nodes to vibrate their finger rings if they think they are in a conversation with the originating Intermediary.

All these buttons and LEDs, however, are not necessary for the software's main functionality as a hub between the sensor network and the intermediary. Communication happens via TCP/IP socket messages. The sensor network hub is a server, and can accept multiple clients that can each send messages to the hub.

Each arriving message is equivalent to a button press on the graphical user interface. If the message is a request, the answer from the nodes is sent back over the same socket to the Intermediary.

For example, upon an incoming call, the Intermediary queries its own Conversation Finder node about the conversation size of its user, and the talk-to-listen ratio. Each request is sent as an ASCII string via the socket interface to the hub, e.g.:

```
REQUEST_CONVERSATION_SIZE,4
```

This is asking Conversation Finder node 4 about its ongoing conversations, and will result in an answer like:

```
CONVERSATION_SIZE,4,2
```

This means that Conversation Finder node 4 has two conversational partners.

If the Intermediary sends the message:

```
REQUEST_TALKLISTEN_RATIO,7
```

it may get back an answer like:

```
TALKLISTEN_RATIO,7,76,23
```

This means that node 7 reports that its user has been talking during 76% of the last minute, and during 23% of the last 15 minutes.

In the next section, I will summarize all possible messages that can get sent between Intermediary, Conversation Finder nodes, and Finger Ring nodes.

4.8.3. Sensor network messages

As mentioned earlier, the protocol has a 4-bit message space and a 4-bit ID space, which allows for 16 different kinds of commands, as well as 16 different node identities (IDs).

The following table shows all implemented messages. There are four types which differ depending who sends them and who reads them:

1. Intermediary – Conversation Finder
2. Conversation Finder – Finger Ring
3. Finger Ring - Intermediary
4. Conversation Finder - Conversation Finder

Intermediary - Conversation Finder

Command ID (hex)	*Description*	*Message size (bytes)*	*ASCII command with parameters*
0xE	Intermediary says: If you are in a conversation with me, contract your Finger Ring!	1	CONTRACT_IF_IN_CONVERSATION, Agent ID
0xC	Intermediary says: I request list dump	1	REQUEST_DUMP, Conversation Finder ID
0xD	Conversation Finder says: I send list dump	36	LIST_DUMP, Conversation Finder ID, data1, data2, data3… data34
0x6	Intermediary says: How many people in your conversation?	1	REQUEST_CONVERSATION_SIZE, Conversation Finder ID
0x5	Conversation Finder: I send # of people in my conversation	2	CONVERSATION_SIZE, Conversation Finder ID, amount
0x4	Intermediary says: What is your talk/listen ratio?	1	REQUEST_TALKLISTEN_RATIO, Conversation Finder ID
0x3	Conversation Finder says: I send my talk/listen ratio	2	TALKLISTEN_RATIO, Conversation Finder ID, ratio1, ratio2

Conversation Finder - Finger Ring

0xF	Conversation Finder: contract!	2	CONTRACT, Finger Ring ID, Interruption ID

Finger Ring - Intermediary

0x9	Finger Ring: veto!	1	VETO, Interruption ID
0x8	Finger Ring: veto to own agent	1	VETO_OWN, Finger Ring ID
0x7	Finger Ring: send heartbeat	1	RING_HEARTBEAT, Finger Ring ID

Conversation Finder - Conversation Finder

0xA	Conversation Finder: I send talk message	1	TALK, Conversation Finder ID
0xB	Conversation Finder: I send heartbeat message	1	HEARTBEAT, Conversation Finder ID

4.9. Issue Detection

This section describes the implementation of a specific sub-system of the Intermediary, the Issue detection infrastructure.

One part of the Issue Detection infrastructure is a set of PERL scripts that continuously every hour captures bags of words from the user's sent mail (separately for message body, quoted text, subject lines, to lines, going through the user's IMAP sent-mail folder as an robotic mail client), ToDo list (web based), and the user's Google web search strings (via modified API).

The system also harvests once a day a bag of words from the user's home pages, for capturing long-term interests. During all harvesting processes, a stop list with the most common 10,000 words is used.

In addition to the speech recognition server, another piece of software matches the bags of words with the speech recognition output, and returns what it thinks this call is about, and how important this is to the user, by showing the importance levels of the matches it found. Importance for ToDo list entries decay the further down they are in the list. Web searches and sent email message have decaying importance: the further in the past the events are, the less importance they get assigned (subject lines decays slower than message body, though, since they are more concise).

In order to go beyond simple literal word matching, a more sophisticated mapping is needed, such as 'fuzzy inferences' between what the caller says and the bags of words. (None of the following options are implemented yet.)

One option may be to expand the existing bags of words with synonyms from WordNet (Miller, 1995) [138], so that "dinner" will match "supper," etc. The right sense of a word could be guessed from the words of the context. Another option may be using the *Openmind* corpus (Singh, 2002) [192] that returns bits of common sense for a word, or even an expression—something WordNet can't. Yet another one is using *OMCSNet* (Liu et al., 2004) [117], a semantic network that is mined from Openmind, but has very clear relationships between the concepts ("is a", "is part of", etc).

All these fuzzy inference mechanisms would go beyond what CLUES filtering (Marx et al., 1996) [128] is capable of. At the same time, they also increase the bags of words. The speech recognition engine is provided with the bags of words as a dynamic vocabulary (XML file), so that it is more likely to recognize them if they would occur during the conversation.

The resulting percentages are then added up, so the Intermediary doesn't look at just one word, but the compound 'relevance' of the recognized words.

5. Evaluation

5.1. Overview

5.1.1. Approach

Evaluating this thesis work is a challenge that is entirely separate from its technical implementation. Evaluation is non-trivial because the current system may require fundamental *paradigm changes* in order to get accepted and taken seriously by users.

Paradigm changes do not happen over night, and it is clear that nowadays people (a) will have problems with not being in charge of their own cellphone settings, (b) will get interrupted more intrusively by a small animatronic device than by cellphone ringing or vibration, and (c) will feel awkward talking *to* their mobile devices instead of *through* them, just to name a few concepts used in this thesis work. But it is likely that there will be *social adaptation* that will work in favor of these novel concepts, and that people in 5 to 10 years will be more likely accept the very paradigms that seem very strange to them now.

One cannot predict paradigm changes. However, it is possible to *plant the seed* to such paradigm changes. This thesis tries to do exactly that. I am convinced that only by showing people with real existing prototypes what *could* be, it may be possible to predict what *might* eventually get adopted and become common.

An evaluation is about how people react to novel concepts and technologies. There are several ways to measure that.

5.1.2. Methods

This evaluation makes use of three different methods:

- Quantitative behavior measurements (if possible)
- Questionnaires and follow-up semi-structured interviews
- Semantic differential measurements

Quantitative behavior measurements and questionnaires will be described in detail in the respective user study sections. The semantic differential method will be introduced in the following section.

Semantic Differential method
Rather than measuring the "efficiency" of a user interface, which is not what this work is about, a different approach would be to measure *attitude changes*. The rationale behind this evaluation is the following: The question is not, "Is user interface A more efficient than B?" but rather: "How does the user's attitude change after being exposed to some key aspects of this thesis work?" or "How does the users' attitude towards this novel technology differ from her attitude towards traditional technology?"

A thoroughly researched and widely accepted instrument for this purpose is Osgood's **Semantic Differential** (Osgood et al. 1957) [152], a methodology for quantifying connotative semantic meaning. The advantage of using this specific psychological instrument is that it is known very well how to interpret the results, where its limits are, etc. This is much less likely the case with newly developed instruments. In addition, there is a large number verification and follow up studies certifying validity and reliability.

A semantic differential, in its simplest form, is a method for measuring a participant's attitude towards an artifact or concept. Often measurements are taken before and after a participant has been exposed to the artifact or concept, or letting participants compare two concepts, allowing the researchers to measure a possible attitude change or difference—hence the term Semantic Differential. Such a differential can be interpreted with higher validity than an un-calibrated single measurement.

A semantic differential is implemented as a series of attitude scales. Each participant is asked to rate a given concept/artifact on a series of bipolar rating scales, such as 'angular – rounded', 'weak – strong', 'tense – relaxed'.

In addition to a 17-dimensional concept space, subgroups of the scales can be summed up to yield scores that are interpreted as indicating the individual's position on three underlying dimensions of attitude toward the concept/artifact being rated. These dimensions are determined using factor-analytic procedures. In many cases, it has been found that the three highest loading factors stand for the following three dimensions of attitude (Kidder, 1981) [97]:

- Evaluation (E): the individual's evaluation of the object or concept being rated, corresponding to the 'favorable-unfavorable' dimension in more traditional attitude scales
- Potency (P): the individual's perception of the potency or power of the object or concept
- Activity (A): the individual's perception of the activity of the object or concept

However, in order to conduct this data reduction in a meaningful way, the number of participants has to be least 5 times the number of scales. This is not the case in these user studies, so comparison was done only on the scales level.

5.1.3. Hypotheses

Although it is not possible to evaluate all elements of this thesis work, there are two specific hypotheses that are at the core of this thesis and can be tested systematically in a formal setting:

Hypothesis 1:
If participants are given the means to anonymously veto upcoming cellphone interruptions by responding to a subtle pre-alert in the form of slight vibration on their finger ring, they will veto more during group-focused settings than during non-group focused settings.

Hypothesis 2:
Interruptions of a social setting caused by human style non-verbal alerts (gaze, posture), which qualify as both subtle and public, are perceived by bystanders as less annoying and less intrusive than interruptions by a ringing phone.

5.1.4. Target population

To contrast with the feedback received from Media Lab students as well as visitors of the Media Lab, goal was to recruit study participants who are neither Media Lab students nor Media Lab sponsors. Obviously, the more diverse the sample is, the more valid the evaluation results will be. For practical reasons, focus was put on administrative personnel and staff, as well as students and faculty from other departments.

The first user study conducted (section 5.2) addresses hypothesis 1, the second user study (section 5.3) addresses hypothesis 2.

5.2. User study on Social Polling

If participants are given the means to anonymously veto upcoming cellphone interruptions by responding to a subtle pre-alert in the form of slight vibration on their finger ring, will they distinguish between different social settings? Will they more likely disallow interruptions in a cognitively demanding group-focused setting, and will they more likely allow interruptions from cellphones during ‘group downtime’? Will a majority of the participants implicitly agree on when it is appropriate to get interruptions, and when it is not?

5.2.1. Pilot study

The pilot study was conducted to obtain information about parameter thresholds. Different vibration patterns were tested, and it was determined that a single vibration burst of one second on a participant’s finger is perceivable yet not disruptive. Furthermore, it turned out that the ratio of collective award vs. individual award during the game (see below) is 1 to 10 in order to balance the behavioral motives. Because some participants had to suppress the reflex to press the ring switch when it vibrates, as in ‘picking up a call,’ a trial run for the game and ample try-out time was scheduled for the user study.

5.2.2. Experimental procedure

The 45-minute user study involved a simple card game. One experimenter distributed a deck of cards to a group of three participants. Then the cards had to be put down in a specific order, one by one, on a single pile in the middle of the table. Each game lasted 70 seconds, and a clearly visible clock showed the count down. The more cards the group could lay down, the more money each participant earned. For each card on the table, each participant received 5 cents. There were multiple games per session. In between the games, there were pauses for reshuffling and redistribution of the cards. Although the game was simple, it required the full attention of all participants; the pauses in between, instead, were low stress periods.

During the whole session—both during the games and the pauses—participants received short phone calls by a remote experimenter. These calls allowed the participants to earn additional money: they were asked a simple question ("What is 13 times 7?"), and if the participant—and only the participant on the phone—answered correctly, he or she received a 50-cent bonus.

Participants were given subtle pre-alerts in the form of a short vibration of their finger ring when any call came in (not just for their own cellphone). Each participant then had the chance to veto it anonymously by pressing the micro switch on his or her finger ring. Every participant was given the same pre-alerts, and at the time of the pre-alert no one knew who would get the call. The ultimate goal of the game was to earn as much money as possible, either from collective or individual rewards; deciding on which to focus was up to the participants.

All sessions were videotaped (Figure 2) with multiple cameras and transcribed to obtain exact timestamps of all events (pre-alerts, calls, vetoes, etc.) In addition to the transcripts, all participants filled out pre and post study surveys (semantic differentials with 17 bipolar scales).

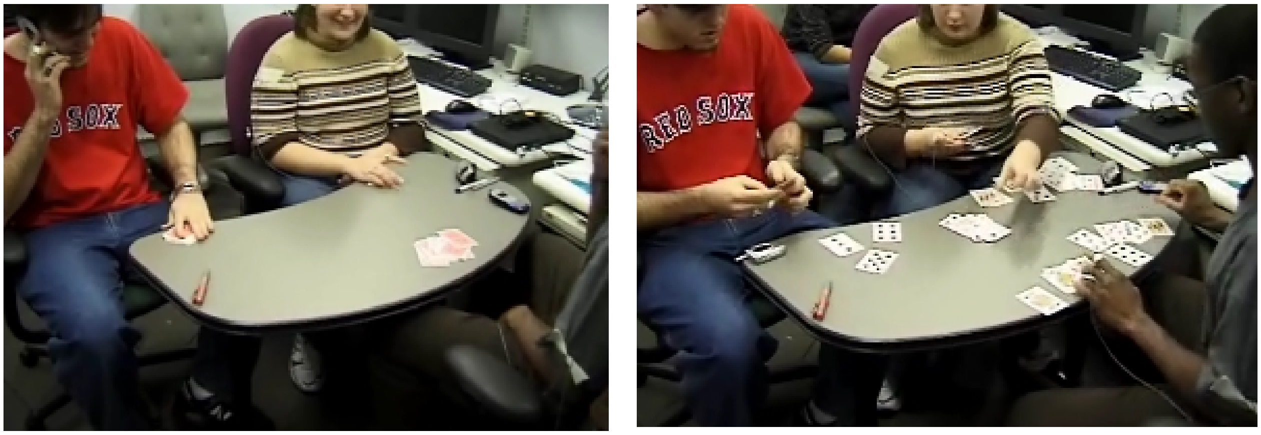


Figure 64: Relaxed during pause (left), highly concentrated during game (right)

5.2.3. Results and discussion

The study consisted of two group sessions, each with three participants. In total, 30 pre-alerts were issued, 15 during the card games and 15 during off-times. The total length of games and pauses were equal. All vetoes across all groups were added up per setting (during card game, during off-times). In all but one case, there was either no veto or one veto per pre-alert.

As Table 1 shows, vetoes happened more than twice as often (53%) in the high attention, collective activity setting than in the 'Pause' setting (20%). Even with our relatively low N, the mean differences between the two settings became statistically significant (p=0.05, t(28) = 1.70, single-tailed t-test): the participants indeed vetoed more during the games than during the pauses.

Setting	*Pre-alerts issued*	*Vetoes received*
Group game	15	8 (53%)
Pause	15	3 (20%)

Table 1: Results

During the de-briefing, one participant voiced concerns that "random people, like in the bus, could disable my phone." The Conversation Finder nodes, which guarantee that only people in the same conversation with the user can veto and not just any person close by, were not necessary for this study, so the participants did not know about it.

Another participant objected that other people might (accidentally) veto important calls, e.g., from a hospital. It was explained to her that a co-located person's veto is just one input of several for our conversational agent that converses with the caller, and is trying to recognize emergency keywords such as "hospital," "accident," etc. at any point in the conversation, and would override vetoes.

Although the current experimental design is based on an egalitarian approach, variations might be worth exploring: e.g., all participants of a conversation are alerted and allowed to veto except the user who owns the interrupting device; more than one veto is necessary to avoid an interruption (majority approach); different users have different weights in the vetoing process (which would require the identity of the vetoers to be disclosed).

The following section describes the second user study, which addresses the second hypothesis: Interruptions of a social setting caused by human style non-verbal alerts (gaze, posture), which qualify as both subtle and public, are perceived by bystanders as less annoying and less intrusive than interruptions by a ringing phone.

5.3. User study on Interruption by Animatronics

In a broader context, this user study asks the question if a physical embodiment of a call handling agent facilitates the mental separation between talking to remote others and co-located people. Due to the novelty of "talking to a stuffed animal," such a claim is currently ludicrous, except perhaps among children. There is, however, ample evidence, based on observations of adoption of mobile telephones and corded and cordless headsets, that people will change the way they converse over a phone.

One can, however, evaluate the claim that an animated embodiment will lead to less discomfort to the co-located third party, especially during the initial transition from local conversation to speaking over the phone. Motivated in part by the methodology of Love et al. (2004) [122], it was decided to interview participants while staging interruptions using both conventional and animatronics telephones. Participants' reactions were examined by observation of their videotaped behavior, a questionnaire based on semantic differentials, and comments in semi-structured post-exposure interviews.

5.3.1. Experimental procedure

Tested were 10 participants, age 25 to 55; 4 were male, 6 female. The participants were administrative or support staff from our building who had little or no previous contact with the project. Each session took about 30 minutes.

First, participants had to be desensitized to the animatronics, so that the novelty factor of the "squirrel phone" would not dominate any other effects. Participants sat facing the interviewer, who was surrounded by five animatronic creatures (our earlier prototypes, a motion-sensing singing bird, a life-like robotic cat, etc.), and the numerous stuffed animals that routinely adorn the interviewers computer monitors. For the first five minutes or so, while participants read and signed the two experimental consent forms, the animatronics were all in motion from time to time. Participants looked at them, and sometimes made comments ("What is this, a zoo?") indicating awareness of the creatures. Then the interviewer pointed out that the squirrel was also a phone, shut down the noisiest of the props, and proceeded with the interview.

While asking questions about participants' use of mobile phones, voice mail, and email, he was twice interrupted by a confederate, over the conventional telephone and the animatronic phone (in random order). The telephone was answered on the second ring. The squirrel phone alerted by "waking up" and looking about. Both devices were used in speakerphone mode, answered in approximately the same amount of time, for a conversation of similar duration. If participants had not noticed the squirrel phone's activity or heard its servos, the interviewer said, "Someone is calling" before squeezing the squirrel's paw and saying "Hello?" The two interrupting phone calls lasted about 30 seconds each, out of a 10-minute interview.

The squirrel was located in between the interviewer and participant. Its default status was ‘asleep,’ that is curled up and breathing slightly. When trying to get attention, it raised its head, opening its eyes, and nervously looking left and right. During the call, it looked straight, moving its head only slightly, blinking occasionally. After the call was done, it fell asleep again. The animatronics’ behaviors were triggered by a confederate who made the phone calls and had a view of the experiment area via CCTV.

A final questionnaire consisted of two semantic differentials and a traditional survey. As described earlier, a semantic differential is a method for quantifying connotative semantic meaning. It measures a participant's attitude towards artifacts or concepts, and is specifically useful to measure the relative difference between two concepts. A participant is asked to rate a given concept on a series of 17 bipolar semantic scales, such as ‘traditional – progressive’, ‘simple – complicated’, etc. She is asked to describe how she feels about a certain concept by placing a check in one of the six spaces between each word pair (similar to a Likert scale). The concepts the participants were asked to rate were:

1. “The ringing phone interruption during this interview”
2. “The squirrel phone interruption during this interview”

In addition to the two semantic differentials, the participants were asked to fill out a short traditional survey and participate in a short semi-structured interview.

5.3.2. Results and Discussion

In the following subsections, the results, of both quantitative and qualitative nature, are discussed.

Quantitative results

The null hypothesis was that attitudes towards interruption would be independent of whether interruption was by a traditional ringing telephone or a moving animatronic device. The data invalidates this hypothesis in several ways. When asked whether they would rather be interrupted by phone or the squirrel, six chose squirrel and four had no preference. Since such direct questions often beg the answer, subjects also rated each device on a six-point “annoyance” Likert scale (1=very annoying, 6=not at all). The squirrel was much less annoying (mean = 5.0) than the phone (3.7). The results were significant (p=0.011, one-tailed t-test).

Perhaps more convincing (because the questions are less direct), we found statistically significant pairwise differences in 8 out of the 17 semantic differential scales (Table 2, p=0.05, two-tailed t-test).

		Phone mean	*Squirrel mean*	*p*
traditional	*progressive*	2.0	4.5	0.002**
friendly	unfriendly	3.9	2.5	0.029*
serious	*humorous*	3.7	5.2	0.021*
stale	*fresh*	2.2	5.1	0.00003**
work	*fun*	1.6	4.9	0.0002**
relaxed	tense	3.7	2.3	0.0498*
bright	dull	4.3	2.7	0.0406*
masculine	*feminine*	2.2	3.7	0.0183*

Table 2: Significant pairwise differences; scale values: 1-6

When participants compare the interruption by a ringing phone with the waking up squirrel, they rate the squirrel significantly more **progressive, friendly, humorous, fresh, fun, relaxed, bright**, and **feminine**. (An EPA analysis was not attempted because of too low N.)

These findings come from semantic differentials that measure the *connotative meaning* of a concept, as opposed to its *denotative meaning*—the difference being that the measured attitudes are rather emotional than rational.

This means that even though participants, if asked directly to chose between ringing phone and animatronics interruption, may not consistently prefer one over the other, their *affective* attitudes towards the two choices still differ significantly and consistently. This can be interpreted as follows: the proposed novel technique for alerting and phone interruption will be perceived differently from traditional alerting techniques mainly on an emotional level.

There were no statistically significant differences due to gender or recency—i.e., the most recently experienced interruption was not more annoying.

Qualitative results

Generally, the participants grasped well the function of the animatronics. When asked to describe it, one participant said, “It is a stuffed squirrel that is kind of animated, and the squirrel would sit and kind of doze off until the phone rang, at which point the squirrel would wake up and its eyes would open, and by just touching its paw he then could talk to the phone by talking to the squirrel.”

Overall reactions were quite positive. “It amused me... I didn't mind it at all.” ”I like it. I wouldn't mind one in my house.” “I think it is cool—I want one.“ “Pardon me for using the word: it's kinda goofy in a way that I really like.”

If asked about its intrusiveness: “I find it lot less objection-able [than ringing].” “It's the cutest... it's cute! I dunno, say it's a fuzzy little... different way, I mean phones are so... sterile, I hate ringing phones, blaring phones!” “The phone ringing is definitively much more invasive that what this [animatronics] is doing. I do think it would be less invasive to the conversation what this was doing than even just a ringing phone—even if he decided not to pick up."

The efforts to desensitize the participants seemed successful. One participant noted that the animatronics activity in the office "was like background. It's like when you have the TV on—background noise." Another one said, "I noticed that there were other animatronics, making little sounds and moving around, but I quickly tuned them out. I don't know if they stopped moving... When we started talking I tuned them all out, pretty easily."

One participant noted that ringing is an interruption mode that masks all other audio—it's an exclusive block on all other activity in the channel, even before the call is answered. Indeed, subjects tended to shift their gaze to the ringing phone much more than the squirrel, and usually stopped speaking as well.

Some participants compared the sound of the servos with the *sound* of a cellphone vibration alarm. One mentioned that he is sensitized to this sound, so immediately guessed that the sound of the waking-up squirrel meant an incoming call; the motors that make cellphones vibrate are indeed very similar to the motors that make the squirrel move.

During de-briefing, about half of the participants reported that they did not notice the squirrel waking up. This suggests that moving animatronics would not adversely affect co-located people—and a priori be more socially intrusive than a traditional phone—and contradicts a common concern expressed about this work. It may also indicate that the squirrel's alerting behavior was a bit too subtle; perhaps it should also make a chattering sound when a call comes.

Despite the small sample size, reactions to the phone and squirrel conditions were so different that the study quickly delivered statistically significant results, and for that reason not more subjects were tested. Since it is based on a large number of dimensions, a complete semantic differential would have required very many more subjects. The subjects' comments and the analysis of their reactions both by the interviewer and later on videotape were rich; the quotes above are representative but a small fraction of the total.

There were, however, some limitations or reservations by the subjects, mostly around the particular animal forms chosen, and clearly some sensitivity to the sounds made by some of the desensitizing props (which were active mostly while subjects read consent forms). For example, referring to the rather loud robotic cat, one subject said: "I am not even sure if the squirrel does it for me, but I'd take it over the cat. If that cat meowed like that all the time, I'd kill it..." A related theme was that subjects clearly had strong preferences for different kinds of animals. And some realized that simply hiding the phone doesn't solve all its problems. "I don't think it makes the cellphone any less offensive in offensive situations." And from another subject: "Just because it is dressed up as a cute squirrel doesn't mean, in a restaurant and somebody's squirrel rings, it will be just as annoying... It might cause an accident if somebody drives by and sees you talking to a squirrel."

But this same subject also noted: “It's subtle—it's not jumping up and down, making lots of noises—it's just there.”

Animacy vs. real human-style non-verbal cues

In order to alert people, the squirrel in this experiment wakes up and looks around. It does not have the sensing capabilities to make actual eye contact with the people around it, nor does it know when people are looking at it. As studies with Kismet show, making eye contact is profoundly significant in people feeling if they were talking to a social presence (Breazeal, 2002) [19].

Although the squirrel does make use of any eye contact sensor and is not aware of the location of the user’s head, some of the embodiments suggested earlier would be able to do so if they were worn or mounted at a known location on the user’s body. E.g., a bird on the right shoulder of the user knows that by turning its head to the left side it will face the user. Similarly, a creature in the user’s chest pocket knows that by looking upwards it will look towards the user’s face. But even when the relative location of the intermediary allows it to know when it is facing the user, it would still need an eye contact sensor to register eye contact.

The squirrel also does not use human-style attention getting gestures per se, since it is not able to wave with its hands/paws. What it really conveys is animacy and alertness, expressed as waking up and looking around.

However, it appears that at least for this interruption study, using human non-verbal cues wasn't critical, although perhaps people would have noticed the squirrel more often than 60%.

In summary, the behavior of the squirrel in this study was merely inspired by human non-verbal cues. What really matters, however, is finding a solution that handles interruptions better than a conventional telephone.

Some questions remain. When are non-human cues sufficient? When are animacy cues enough, and when is it important that they actually become more human-like rather than just conveying animacy?

There are two answers to these questions.

First, one has to take into account the amount of information that the embodiment intends to convey. If its only purpose is to wake up and convey the importance of the interruption, it can be done easily via the speed and abruptness of the movements, by varying the openness and blinking frequency of the eyes, etc. Without actually using true eye contact and waving gestures, this simple behavior will clearly convey a level of nervousness and alertness that will be mapped intuitively by user and bystanders to the importance of the interruption. It has to be noted that although the embodiment does not know when the user looks at it, it is still aware of the user’s tactile actions (switches are located in the animatronics embodiment). This information is used to determine simple positive and negative acknowledgements from the user’s side.

The semantic meaning of these behaviors varies depending on the status of the conversational agent's finite state machine.

Second, if the embodiment needs to know if the user has noticed it—for example when using escalating alerting schemes, then it needs more sophisticated sensors to detect eye contact, and may need more human-like gestures (waving) in case that merely looking around does not attract the user's attention enough or in time.

In short, the more interactive the embodiment is intended to be, the more it needs additional sensors and should use human-like gestures instead of mere displays of animacy and alertness.

Human style interactivity, however, was not the goal of the current implementation of the Intermediary, since the system is not intended for truly human style interactions.

Although no eye contact sensor is available, the animatronics has some simple interactive behaviors since it can 'hear' the user and employs pause detection for human-style turn-taking behaviors, as well as registers the user's touches on the animatronics extremities. Still, these mechanisms can obviously not compete with truly human-style interactive behaviors. The main focus of the Intermediary embodiment was rather to use human-style non-verbal cues for alerting and interruption.

How are my squirrel's cues of animacy and alertness different from alerts by peripheral displays such as Ambient Device's *Orb*?

On one hand, character embodiments may be more intuitive than simple color combinations: although people can learn the meaning of simple color LEDs relatively fast, these combinations are not intuitive (Campbell et al., 2004) [26] Furthermore, a robotic user interface shares the same physical three-dimensional space with user and co-located people.

But most importantly, the difference between an ambient display and an embodiment is that the embodiment has likely a different, if not just more 'appeal.' As described earlier, cuteness is only one aspect of appeal, but it appears that zoomorphic and anthropomorphic embodiments may easily trigger emotional reactions on the visceral, behavioral, as well as reactive level of design, whereas ambient displays may trigger emotional reactions mainly on a behavioral level (Norman, 2004) [150].

6. Related Work

The related work section is split in several subsections to cluster the references. First, there are three subsections that address work related to Autonomous Interactive Intermediaries as a system:

- Mobile communication (section 6.1)
- Conversational agents (section 6.2)
- Socially intelligent agents (section 6.3)

Then there are subsections that address specific parts of an Autonomous Interactive Intermediary, such as:

- Conversation Finder and Finger Ring (section 6.4)
- Issue Detection (section 6.5)
- Animatronics and embodiment (section 6.6)

6.1. Mobile communication

6.1.1. Social impact of mobile communication

This thesis work addresses issues of *social impact of mobile communication*. The impact that mobile communication has on our lives seems intuitively clear to the users of mobile communication, as shown in rich ethnographic studies by Plant (2000) [157] and Rheingold (2002) [167]. The former provides a wealth of behavioral observations in the domain of cellphone use, the latter focuses on the effects mobile communication has on society and groups. Anthropological research determines the social impact of mobile communication upon specific age groups, like adolescents, where it has become the primary mode for socializing (Blinkoff, 2003) [16].

Social impact of mobile communication is also addressed by work describing the 'etiquette' of cellphone use, like Ling (1997) [113] and Laufer (1999) [105]. The former examines how people deal with inappropriate mobile telephones use, drawing heavily on Goffman's notion of drama and staging (e.g., Goffman, 1966) [67]. The latter lightheartedly shows up "the right way," "the wrong way," "the new way" and other ways of using a cellphone, coming to the insight that "the rapid growth of wireless technology is overtaking society's ability to accommodate the saturation of mobile phones in our midst without some common guidance." (Laufer, 1999) [105] This thesis work tries to alleviate this situation by adding an active Intermediary that will have some limited sense about what behavior is socially appropriate.

However, the social impact of mobile communication is still not very well studied systematically by researchers. The most comprehensive sociological survey of mobile communication impact is done by Hans Geser (2002) [66]. He carefully analyzes the impact of mobile communication on five different levels: implications for human individuals, on the level of interpersonal interaction, implications for face-to-face gatherings, consequences on the meso-level of groups, and implications on the macro-level of societal institutions. Especially his ‘dissection’ of the third level, impact on face-to-face gatherings, is highly relevant to this thesis work.

“The unpredictable, uneasy intrusion of distant other (…) strains the capacity of individuals to switch roles an to redirect attention very rapidly at any unforeseen moment, a well-known source of tiring psychological stress.” (Geser, 2002) [66] He lists four highly negative, destabilizing influences of cell phones on ongoing face-to-face interactions:

1. Calls occur at unpredictable times, and therefore cannot be anticipated and integrated into the local discourse.
2. Deeply anchored norms and habits demand that calls are answered at the moment they come in, so that the local interactions are disrupted even at highly critical moments.
3. When we answer a cellphone call, we are getting involved in a bilateral communication process that is completely segregated from the local interaction and is highly opaque to bystanders. The called party can either leave the place of co-local interaction, or stay and suspend current activities (leaving bystanders helplessly for an undefined time), or take the call and at the same time keep current activities ongoing (which does not work when they consist of verbal communication). All these behaviors are highly disruptive.
4. Even worse, there seems to be a deeply rooted habit to focus attention completely on the communication with the caller, therefore disengaging oneself psychologically from the face-to-face discourse at least on the verbal communication level.

Geser writes, “The demand for social control will rise because, in a world where social differentiation can no longer be based on spatial segregation, it has to be increasingly secured by controlling individual behavior. Such control can be realized in three forms:

- *Intraindividual self-control*: e.g., in the case of users avoiding or shortening incoming calls in order to concentrate on ongoing collocal interactions
- *Informal interindividual group control*: e.g., in the case of collocal partners showing impatience when cell phone calls go on for longer than expected
- *Formal institutional control*: e.g., in the form of regulations prohibiting cell phone calls during school or working hours.”

These social control mechanisms might be necessary for society, but will be perceived as highly intrusive to the individual. This thesis work tries to find another way, exploring an alternate solution that makes the above mentioned control mechanisms obsolete.

Although Haddon (2000) [75] claims that the ‘friction between mobile users and co-present others’ (Cooper, 2001) [32] has been noted by a range of observers and is well documented in both qualitative research (Ling, 1997) [113] and in quantitative surveys, there still is no satisfying technical solution to this problem.

There is a second strand of research that touches on the social implications of mobile communication: the social aspects of *mobile computing*. Although not directed specifically towards mobile communication, the findings of mobile computing research are still applicable. E.g., Dryer et al. (1999) [44] investigate the social impact of mobile computational devices that are designed to be used in the presence of other people. These devices may promote or inhibit social relationships. They consider four social relationships: interpersonal relationship among co-located persons, human-machine relationship, machine mediated human-human relationship, and relationship with a community. Dryer et al. refer to Social Computing as the interplay between a person's social behavior and her interactions with computing technologies. They draw on research from social interfaces (starting from the fact that humans can react socially to artifacts, and pervasive computing will lead to proliferation of artificial social actors), computer-supported cooperative work (CSCW), interpersonal psychology, and community research. The authors conduct several extensive lab experiments to correlate these factors, comparing laptop, PDA, belt-worn wearable, and wearable with head-mounted display. The most important finding is that because these devices have not been designed to support social interactions, they can make users appear socially unattractive. Therefore, the authors suggest a 14-item "checklist for social computing," for devices that are designed to be used in the presence of other persons. This list describes factors that they expect to have an effect on interaction outcomes, and includes items such as disruption (does the device disrupt individuals' natural social behaviors, such as referring to shared information while interacting?), perceiver distraction (does using the device create a distraction for nonusers?), and user distraction (does the device place a high cognitive load on the user during use?)

These findings will become more and more relevant to the social impact of mobile communication devices because of the ongoing convergence between PDAs and cellphones, and will aggravate the friction between mobile communication users and bystanders.

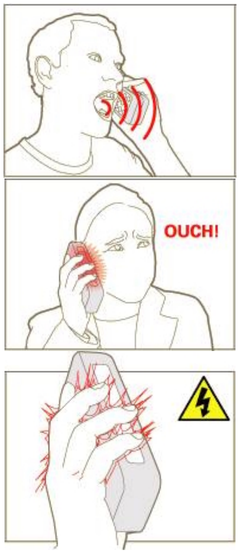


Figure 65: IDEO's Social Mobile #1

Artists and designers have also tried to come up with solutions to the negative social influences of cellphones on bystanders. E.g., IDEO's case study Social Mobiles [86] describes "phones that in different ways modify the user's behavior to make it less disruptive." One version is a phone that delivers a slight electric shock (Figure 65) depending on how loudly the person at the other end is speaking. As a result, the designers hope the two parties are induced to speak more quietly.

6.1.2. Context-aware mobile communication

The research domain of context-aware mobile communication—a sub domain of context-aware systems (e.g., Lieberman et al. 2000 [110], Selker et al. 2000 [188])—is trying to alleviate the negative impacts of mobile communication by creating systems that use context to facilitate the use of mobile communication devices.

In the mobile communication domain, context-awareness means that the caller can preview the social context of the called person. This could include information about where the communication partner is, or how open and/or available she is to communication attempts.

On a coarse level, one can distinguish between active and passive context-awareness, as well as personalization. In the domain of mobile communication, the first means that the phone automatically adjusts the user's context, based on physical or logical sensor information; the second means that the user changes her presence information on her phone manually, based on suggestions from the phone; the third means that the user manually sets profiles for herself, the way most phones work today.

Avrahami et al. (2003) [3] provide an extended categorization. They distinguish between five solutions:

1. *Single rule solutions*: manual call settings, valid for all calls.
2. Manual filtering solutions: user considers caller ID in order to decide to take a call or not
3. *Multiple rules solutions*: user sets up different profiles in advance for different activities, locations, and people. She has to anticipate and categorize situations in advance.
4. *Automatic solutions*: the system infers the user's context from physical and logical sensor information, and changes the settings (phone off, set to certain profile, or redirect caller to a different medium)
5. *Caller-based solutions*: the user provides the caller with contextual information, either manually or automatically. The decision about what action to take stays with the caller.

The last two categories are of specific interest to this work.

A simple example for a passive caller-based system is *context-call* by Schmidt et al. (2000) [183] (and Schmidt et al. 2001) [182], who suggest making mobile telephony context-aware by exchanging information before initiating a call. Their system, implemented in WAP, is trying to alleviate the situation that the one who uses the mobile phone is responsible to set the phone in a mode that is appropriate for the situation he or she is in. In most cases, this is a binary decision—switch the phone off or leave it on. This results in a trade-off of not being disturbed but possibly missing a call, versus not missing any calls but being possibly unnecessarily disturbed. Although the authors emphasize that people want to be in control about their visibility, and want to share information selectively (which is not implemented in their system), the biggest concern with a passive context-aware system like *context-call* would be that users will not update their status on a regular basis, if they have to do is manually. The authors refer to this problem and suggest an active context-aware approach with automatic context recognition and selection (Schmidt et al., 1999) [181].

There are other, more complex caller-based systems that are similar to *context-call*, e.g., *Live Address Book* (Milewski et al., 2000) [136], *ConNexus and Awarenex* (Tang et al., 2001) [197], *Hubbub* (Isaacs et al., 2002) [88] *Calls.calm* (Pedersen, 2001) [154]. Most of them use a combination of physical and logical sensors to determine the user's context, as suggested by Schmidt et al. (1999) [181]. According to

Milewski et al. (2000) [136], however, updating presence and availability information might be best done by a combination of automatic detection and manual updating, which means a combination of active and passive context-awareness. Barkhuus (2003) [5] found that context-aware applications, especially the active versions, are preferred over the personalization oriented ones, even if the user has to give up partial control

The above-mentioned context-aware systems deliver context information back to the calling party, which may or may not take this information into account. However, these systems do not modify the alerting of the called party in any sophisticated way, which would be a different aspect of context sensitivity.

The most important work that influenced this thesis is Hansson et al.'s (2001) [77] findings on *subtle but public alerting*. The authors discuss the design space of notification cues for mobile devices, and propose an exploration of the space that combines the two dimensions of *subtlety* and *publicity*. They suggest combining the properties of subtlety and publicity when designing notification cues in order to make them fit more smoothly into social settings. Public and subtle cues are visible to co-located persons, and can therefore avoid unexplained activity.

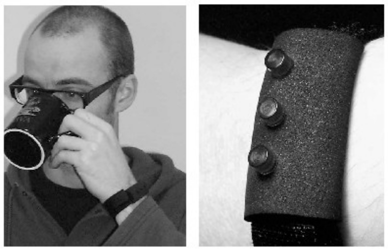


Figure 66: Reminder bracelet

This results in user interfaces for mobile devices that support (and encourage) public but subtle alerting schemes. An example of a crude subtle/public mobile communication alert would be a pager emitting a very short, low volume beep. If designed correctly, it might be unobtrusive enough not to disturb the social environment, but still audible enough to be public. Much more sophisticated, however, is the *Reminder Bracelet* (Figure 66), a notification tool that is worn on the wrist and connected to a phone or PDA. It notifies the user in the periphery of her attention of scheduled events in a subtle and silent manner using light, color, and patterns (Hansson et al., 2000) [76]. It is deliberately designed so that not only the user can see the alert, but also co-located persons. One could ask why not just use the vibration alarm which is built into many phones already. Although such tactile displays are private, non-intrusive and silent, there are some major differences to the Reminder Bracelet. A vibrating device is not visible to co-located persons, and it is therefore hard for others to understand why, for instance, the user suddenly leaves from a meeting. "It provides the user with completely private information and therefore it has a low degree of publicity. An audible signal has a high degree of publicity, whereas a device such as the Reminder Bracelet falls somewhere in between these two extreme cases. Using notification cues with a higher degree of publicity allows other people present to interpret the situation at hand, e.g., in terms of causality." (Hansson et al., 2000) [76].

However, there are limitations in the usefulness of subtle/public alerts in the mobile communication setting. Hansson et al. seem to come from the assumption that an alert that explains our behavior is a good alert: if we get an alert, we should be excused to interrupt our current activity (e.g., interaction) and do something else. That is not something new: In the most basic sense, if we interact with somebody, and suddenly an internal "alert" goes off ("something comes to my mind", "suddenly I remembered that..."), then we usually try to interrupt our current activity gracefully and politely, and switch to the new behavior. It is not clear if appropriate subtle but public alerts can take the burden off us, so that we don't have to say politely: "Would you excuse me for a second, something important has come up?" In short, transparency is not equal to acceptability. However, eventually social norms will decide if subtle and public alerts are sufficient to excuse the user; but even if such alerts are insufficient, they are more useful than subtle private alerts, and certainly more appropriate than any kind of intrusive alert.

The Intermediary of this thesis work relies on subtle but public alerting, but in a way that goes beyond what Hansson et al. described: the current system uses non-verbal social cues, like eye gaze, to interrupt the user and collocated people. These signals happen to be both public and subtle, and satisfy Hansson et al. alerting scheme.

Other related research includes Hinckley et al. (2001) [79]. The authors describe sensing techniques to increase the context-awareness of mobile phones, making them more polite and less distracting. They built a PocketPC prototype (simulating a cellphone) that reduces its ring volume as soon as the users touches it, and mutes completely if the user glances a the caller ID and decides not to answer.

The authors also suggest replacing the press of a TALK button (when picking up a call) with a gesture that consists of holding the device, tilting it in a pose typical of talking into it, and detecting the head in close proximity. It is not clear, however, if such a sophisticated sensing mechanism justifies replacing a simple button press, especially when the button is located on the phone so that the thumb of the user touches is anyway when she picks up the phone.

Although these sensing features reduce the amount of ringing, they do not avoid it altogether—which makes it different from the current Intermediary implementation that relies only on non-auditory social cues.

Sawhney et al. (1999 and 2000) [170][171] propose sophisticated techniques for dynamically adapting notification modality and calculating a usage level. *Nomadic Radio* is a wearable system that delivers voice mail (digitized speech) and email and news (text to speech) via auditory channel. This information is played back on two shoulder-worn speakers (Nortel Soundbeam, Figure 67). The alerting begins with low ambient sounds, and then dynamically scales up, going through several levels of intrusiveness, from a subtle auditory cue to full foreground presentation of the message. The scaling up depends on the user's history of usage and the importance of the message being played. If the user has not used the system recently, if she is in the middle of a conversation (as detected via the microphone), or if a message is unimportant, then the system will follow a relatively non-intrusive ramp for outputting information: e.g., it will play a low volume sound of running water, slowly increasing in volume, getting the user's attention, followed by an auditory cue and a short summary. In order to hear the full message, the user has to request it explicitly. However, if the system expects that the user is not busy or if a message is judged to be important, then a faster ambient sound and a full

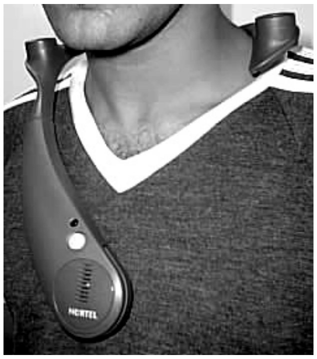


Figure 67: Nomadic Radio

preview will be played. Nomadic Radio also claims to maintain a model of how interruptible the user is at the moment and changes that model based on how often the user overrides the default level of dynamic scaling with which a message is played.

This alerting behavior of Nomadic Radio is certainly more sophisticated than any other mobile communication system. However, having used the system for a few days as a test subject, it became clear to me that having only an audio interface can be perceived as restrictive. Sometimes, the user does not want to use audio modality to interact with an Intermediary, but has no other option. Furthermore, Nomadic Radio is a one-way system that does not allow replying to messages. It is neither duplex nor truly interactive. Of course it also does not use non-verbal cues of any kind.

Other research focuses more on the *cognitive dimension* of attentive interfaces. Horvitz et al. (1999) [84] seek to optimize attention-sensitive alerting and describe a notification architecture that uses probabilistic techniques to balance the context-sensitive costs of deferring alerts with the cost of interruption. It prioritizes notifications, allowing the system to either suppress them or deliver them at an appropriate time, to an appropriate device, using context information such as the user’s calendar or desktop activity.

Although this work is not specifically about mobile communication, it shows an alternative approach to context-aware systems that can be used in alerting on mobile devices.

Roel Vertegaal and his colleagues suggest context-awareness for communication devices on an *attentive dimension* (Shell et al. 2003a [190], Shell et al 2003b [191], Vertegaal et al. 2000 [201], Vertegaal et al. 2002 [202], Vertegaal et al. 2001 [203], Weiss 2003 [204]). Since “eye contact functions as a nonverbal visual signal that peripherally conveys attention without interrupting the verbal auditory channel, (…) humans can achieve a remarkably efficient process of conversational turn taking. (…) To facilitate turn taking between devices and users in a nonintrusive manner, Attentive User Interfaces (AUI) monitor nonverbal attentional channels, such as eye gaze, to determine when, whether, and how to communicate with a user. (…) Since eye movements are not always voluntary, they are best interpreted as an indicator of interest, rather than as a means for control.” (Shell et al. 2003a [190]) (More about AUI in Maglio et al., 2000) [125]

Vertegaal et al. present *EyePliances,* attention-seeking devices that respond to visual attention by the user. The authors use small eye tracker cameras and low-cost EyeContact sensors, mounted on TVs, lamps, but also cellphones, to capture the attention of the user, or rather, when somebody is looking at it. This allows people to use their eyes as pointing devices, and their mouths as keyboards—the so-called *Look-to-Talk* paradigm. The goal for the ‘attentive cell phone‘ specifically is to “implement some of the basic social rules that surround human face-to-face conversations.” (Vertegaal et al. 2002) [202] Currently, the attentive state sets the default notification level on the user’s cellphone (or rather, PocketPC), and allows other users to see this attentive state in a kind of buddy list.

Although these appliances—like the Intermediary—emphasize the importance of gaze as a non-verbal signal, they utilize it in the reverse direction: these researchers use human gaze to signal implicit attention to a computer system, whereas this work is about a robotic interface that uses gaze to alert the user in a subtle non-verbal way. However, a more recent version of the system, called eyePROXY (Figure 68, Jabarin et al. 2003) [90] adds the reverse direction of gaze to the system.

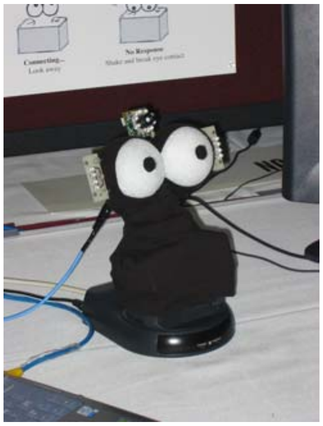


Figure 68: eyePROXY

Although there is no vision system in the current implementation of an Intermediary, it certainly would be an interesting extension to an Intermediary to explore further.

It is obvious that the current Intermediary is very different from the above-described attentive cellphone because an Intermediary—which has a similar kind of attentive sense from the Conversation Finder sub-agents—does not broadcast the user's attentive state, but rather uses this information only in its speech interaction with the calling party. The difference is that there is an autonomous, proactive agent using this information, and the calling party does not have to deal with the meaning of the attentive state of a user before the initiation of the call.

It appears that researchers like Erickson (2001) [56] do not believe that intelligent systems of any kind are capable of interpreting context-aware system's sensor data as well as humans do, so humans have to be kept in the loop all the time. Although these researchers may be correct at present time, I do not share their pessimism for future developments.

There is also research that supports my optimism. Hudson et al. (2003) [85] report a study where they compared statistical models predicting human interruptibility with collected self-report data. It is based on the assumption that a person seeking someone else's attention is normally able to quickly assess how interruptible he or she is. This allows us to behave in socially appropriate and polite manner. A system that automatically and reliably detects our interruptibility could be used to build an intelligent answering machine. Instead of building a system with a vast sensor array, they choose a Wizard of Oz approach to simulate a wide range of plausible sensors, using audio and video recordings. The authors report an overall accuracy of 78% of their predictions for office workers. The authors also discovered that much of that predictive power could be obtained using a single, relatively easy to build sensor that indicates whether anyone in the space is talking—information that the current Intermediary can obtain easily from the Conversation Finder sub-agents.

Hudson et al. believe that using directly observable phenomena is more promising that trying to assess, e.g., an invisible internal cognitive state. However, instead of guessing the user's cognitive state, one could just listen to what the user is talking about.

Eagle et al. (2003a, 2003b, 2003d) [49][50][52] use the content of the user's spoken language to assess his situation. With this novel approach, they try to infer aspects about a user's situation from spoken conversations using contexts and commonsense knowledge. Spoken language recorded on the user's PocketPCs is transcribed; the noisy transcripts are semantically filtered using a commonsense knowledgebase (Singh 2002) [192], and combined with location information. The goal is to accomplish *topic spotting*, or in a wider sense, establishing the conversational situation: the topic and the surrounding context of a conversation. “With only one correct word for every three, even a human would have a difficult time inferring the gist of a transcribed conversation. But just as additional contextual and common sense information can help a human infer the topic of a conversation, this type of information can be equally beneficial to a probabilistic model. Given a commonsense knowledgebase, along with contextual information from these mobile devices, creating a classifier to determine gist of noisy transcriptions becomes tractable.” (Eagle et al., 2003b) [50]

It is obvious that mobile communication devices that ‘understand’ what the user is currently talking about—either on the phone or with bystanders—in the sense that the conversational topic is extracted correctly, would be a very valuable information source for context-aware systems.

6.2. Spoken language conversational agents

Since this thesis work is about an Intermediary for mobile communication, it contains an appropriate conversational agent system in the specific domain of personal assistant for telecommunication. Here is a short overview of related systems.

6.2.1. Interactive conversational agents

Computers as secretaries that are capable of a spoken conversation with humans have always been on the wish list of A.I. Even Licklider et al. (1968) [108] in his seminal paper about the Internet describes them:

> “(...) A very important part of each man’s interaction with his on-line community will be mediated by his OLIVER. The acronym OLIVER honors Oliver Selfridge, originator of the concept. An OLIVER is, or will be when there is one, an “on-line interactive vicarious expediter and responder,” a complex of computer programs and data that resides within the network and acts on behalf of its principal, taking care of many minor matters that do not require his personal attention and buffering him from the demanding world. “You are describing a secretary,” you will say. But no! Secretaries will have OLIVERS.
>
> At your command, your OLIVER will take notes (or refrain from taking notes) on what you do, what you read, what you buy and where you buy it. It will know who your friends are, your mere acquaintances. It will know your value structure, who is prestigious in your eyes, for whom you will do what with what priority, and who can have access to which of your personal files. It will know your organization’s rules pertaining to proprietary information and the government’s rules relating to security classification.
>
> Some parts of your OLIVER program will be common with parts of other people’s OLIVERS; other parts will be custom-made for you, or by you, or

> will have developed idiosyncrasies through "learning" based on its experience in your service." (*The Computer as a Communication Device*, by J.C.R. Licklider and Robert W. Taylor, April 1968)

Licklider does not mention spoken language interactions, but describes many important elements of an Intermediary that has the function of a personalized telecommunication assistant.

Almost two decades later, a captivating vision of an Intermediary with spoken language was presented in Apple Computer's concept video *Knowledge Navigator* (Figure 69, Dubberly et al. 1987, Sculley 1987, Sculley 1989) [45][185][186]: the ultimate assistant avatar in the upper left corner of a tablet computer helps a professor plan his personal and professional life, which includes intercepting and mediating phone calls.

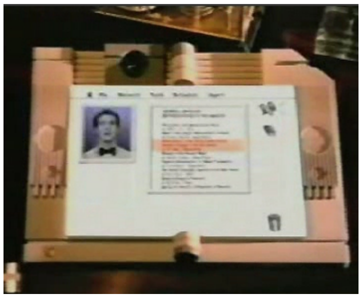


Figure 69: Knowledge Navigator

Conversational agents and voice communication with computers (e.g., Schmandt 1994 [174], Allen 1994 [1], McTear 2002 [133]), including for the telephony domain (e.g., Boyce 2000) [17], have come a long way since Licklider's article—but still have not gotten to the level of the Knowledge Navigator. In the mid eighties, Chris Schmandt conducted very related research on conversational systems:

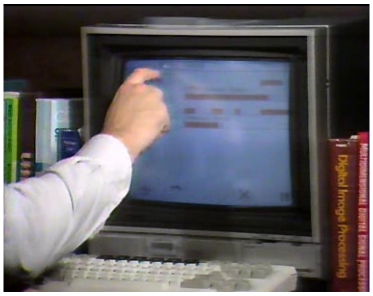


Figure 70: Phone Slave

Phone Slave (Figure 70, Schmandt et al. 1984a, 1984b) [175][176] is an interactive conversational telephone answering machine that exploits people's willingness to participate in a computer driven conversation. It takes voice messages by answering the phone and engaging the caller in a conversation. It plays speech segments and records replies to gather a series of message components, for example: "Who is calling?" "What is this in reference to?" "At what number can you be reached?" "Will you be around?" It's based on the idea that immediately interacting with the caller with a well-timed series of questions increases the likelihood of getting her to leave a message. It encourages the caller to leave a message by guiding the conversation through short questions that invite short responses. Another advantage of this idea is that even without speech recognition, Phone Slave can give an answer to the user's question "Who left a message, and what were they about?" without playing the whole message.

There is also personalization involved on the side of the caller: If a person calls back, the system will recognize her and she will be informed if the owner has heard her message, received a personal reply, etc. Since there is no intention to 'trick' the caller into believing that he is speaking with a person by using a voice with is clearly different from the owner, Phone Slave is a third entity in the communication process, which is focus of the Intermediary work of this research.

However, Phone Slave is a passive system that makes little use of knowledge of other activities its owner is engaged in. The conversational scripts are predefined, which might be bothersome to callers who call often.

The *Conversational Desktop* (Schmandt et al. 1986, 1985) [177][178] is a conversational office assistant that manages personal communications (phone calls, voice mail messages, scheduling, reminders, etc.). Going beyond the functionality of Phone Slave, the Conversational Desktop engages the user in a conversation to resolve ambiguous speech recognition input by applying syntactic and acoustical context to the progress of the conversation. The idea is that the more the system is cognizant of its user's activities, the greater its ability to make correct inferences about its own proper behavior in response to stimuli from the outside world. In order to disambiguate the noisy speech recognition, the dialogue is influenced by the system trying to fill in gaps in a parse tree, e.g., "Schedule a meeting with Chris at <mumble>" will result in a question from the system like "When do you wish to meet with Chris?"

Another element of context-sensitivity is that the system—before playing an audio reminder—first checks the audio level of the microphone, and can postpone the reminder until the user is alone in his office. The general rule is not to interrupt when the user is engaged in some detectable activity. The user can also tell the system "I am away for lunch", and the Conversational desktop will use the outgoing message "out to lunch" from then on. Furthermore, the Conversational Desktop can distinguish if a user addresses it, or somebody else, from its directional microphones.

In the future work section, the authors mention that they wish that the system would take into account the importance of a call when interrupting. Some incoming calls should interrupt at any time, where as most should never interrupt an ongoing conversation. Exactly this problem (among others) is addressed by this thesis work.

The situation now, almost 20 years after Phone Slave, shows that commercializing virtual office assistants with spoken language is still a non-trivial task. Examples of services, some of which have already disappeared, are *TellMe, Portico*, *Nuance Voyager*, *Webley*, *and Wildfire*.

As an example, here's how Wildfire was advertised:

> "Wildfire is a voice-activated, Virtual Assistant that humanizes communications through a simple, intuitive voice interface that gets to know you. "She" provides a single way to manage all of your communications by simply using your voice. She makes calls on your behalf, manages your contacts, takes messages, routes calls, screens incoming calls and even takes and sends faxes, all through intuitive voice commands given directly to Wildfire over any phone."
>
> "(...) Most importantly, Wildfire has a personality. It's not just what "she" says but what "she" does, how "she" does it and how "she" interacts with the subscriber. "She" develops relationships with her users, learning from them and for them, like an actual person."
>
> "Wildfire's adaptive persona evaluates an individual's usage patterns and prompts Wildfire to suggest additional services a user would find valuable, allowing each user to write a "job description" incrementally for his or her own assistant. This benefit allows users start with simple, bite-size pieces of functionality and then, as their comfort levels increase, introduce new services that are appropriate to the way they live and work. Users acquire new services at their own pace and feel in control. They select only those new features they need, then master each set of services before acquiring new ones." (From the Wildfire web page, now offline)

However, the way Wildfire was perceived was slightly different:

> "Wildfire was an early favorite until it started irritating people, industry observers said. (...) It let phone calls actually follow their user whether he or she was in the office, at home, in the car, or somewhere in between. [It] used a software agent to screen incoming calls to glean information and route them appropriately to cell or home phones. The agent appeared in the form of a human (or human-like) voice. (…) Wildfire was really interesting in the beginning, and then it just got annoying. (…) It's one thing to be disrespected by a person. It's another to be disrespected by a robot."
>
> (*Vendors Vie To Fill The Message Gap*, by Barbara Darrow, TechWeb News, September 7, 2000, http://www.techweb.com/wire/story/TWB20000906S0012)

These strong reactions show the fine line that has to be observed when building spoken language assistants: the line between usability and annoyance. The current Intermediary implementation is perceived as less 'annoying' than systems like Wildfire because it specifically addresses the problem of alerting (which makes it less annoying for the user side), and can fall back on social intelligence from the user's surroundings and generic commonsense knowledge (which should make it more acceptable for the calling party).

Other advanced conversational agents are targeted more towards web page interactions, like the Interactive Conversation Interface (ICI) (Gottlieb 1997, 2002) [69][70], demonstrated on the homepage of Jellyvision[5] and summarized in their white paper, is a system that adds a unique colloquial touch to a conversational agent. Gottlieb summarizes the philosophy in a few design principles, which are about maintaining the pacing of the conversation, and creating and maintaining the illusion of awareness for the user, in order to allow the audience to suspend their disbelief. The system itself is a huge state machine, and all the conversational branches are scripted out and pre-recorded by actors.

6.2.2. Voice instant messaging in conversational agents

Voice instant messaging refers to a variation in voice communication where a full-duplex connection, like in a normal phone conversation, is replaced by a half-duplex connection, where only one person can talk at the same time. Therefore, in this variation of voice communication, the user first has to press a button before she can speak. This means that turn taking is technologically resolved, whereas in a full duplex connection the turn taking is socially mediated.

The affordances of this kind of voice instant messaging are similar to well known short-range radio, but is different in that there is not just one channel to which anybody can tune in, but a dedicated one-to-one channel (where one party can be a dedicated group of people) (Woodruff et al., 2003a, 2003b) [206][207].

[5] http://www.jellyvision.com/

The Intermediary has a feature that is unique for a conversational agent: it is able to transform synchronous audio (a caller on the phone) into a sequence of chunks of asynchronous messages between the caller and the user. Or in other words: it can transform a phone call into a voice instant messaging session, acting as a kind of conversational messenger between the two parties.

However, there are systems that do something closely related: they allow the user to play back manually short audio messages to a caller, if he can't or doesn't want to have a synchronous voice interaction.

These passive context-aware systems all share the same goal: they try to lessen the problems of social disturbance caused by cell phone communication by allowing users to respond to telephone conversations without talking aloud. This leads to a novel mixed-mode communication where the caller experiences a voice phone call, and the user something related to a quiet Instant Messaging session.

Taming of the Ring (Pering et al., 2002) [156] works as follows: when the user gets a call, based on caller ID, she can trigger self-recorded messages that vary at the level of promised commitment, e.g.:

- "Hold the line a moment, I'll be right with you." (Hold)
- "I'm in a meeting, leave a message and I'll get back to you in ten minutes or so, press # if this is urgent," (Call Back)
- "I'm in a meeting right now, please leave a message and I'll get back to you soon, press # if this is urgent" (Meeting)

The user can trigger these messages from a watch, which acts as a cellphone remote control. The author writes that the watch form factor was chosen because a user's wrist is a location that can be discretely touched during a meeting without being considered rude.

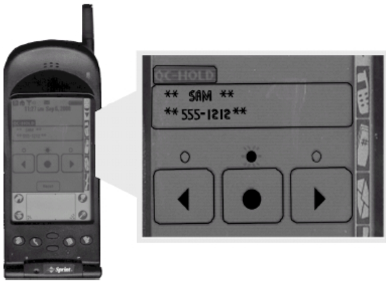


Figure 71: Quiet Calls

Quiet Calls (Nelson et al., 2001) [147] goes a bit further in that the user can not only trigger an initial voice message, but she can also listen to the caller's spoken reactions, and instead of talking, triggering additional voice messages. The user also has only three buttons available (Figure 71), triggering three kinds of messages. Each class of messages supports a 'direction' of the conversation, using a 'Talk-As-Motion' metaphor:

- 'Move into the talk,' engage: Hold the caller while moving to an area suitable for talk
- 'Move out of the call,' disengage: Politely defer a call to a later time
- 'Stay in place,' listen: Listen to the caller without vocalizing

Within each class, several different messages are played back depending on how far the conversation the participants are. This leads to a kind of two-way conversation, since there can be more than one iteration between the caller and the user.

Even further go the commercial In-Call Services by *SoloMio*. Their system intercepts incoming calls and uses text menus to interact with the user, as well as interactive voice responses to communicate with the caller. The user's display is updated continuously to reflect the interaction. Or more in detail:

> "Triggered by an incoming, missed or rejected call, SoloMio sends a dynamic data menu that pops up on the display of the subscriber's handset, giving them relevant and compelling call-handling choices. SoloMio intuitively suggests contextual data menu options, e.g., 'Hold on!' 'Is it urgent?' 'Call you right back…' for the mobile subscriber to communicate in real-time with every phone call. In-Call Service menu options are contextual, driven by network-based personalization techniques. With no configuration needed by the end-user, 'watch and learn' technology adjusts call options automatically - taking into account overall subscriber communication patterns and interaction with individual callers." (SoloMio home page, http://www.solomio.com/)

In-Call Services are different from Quiet Calls in that they allow the user to have more complex respond options. However, more research has to be done to determine if more complexity and more choices make the system better or worse than a clear Engage-Listen-Disengage structure of all possible answers (like Quiet Calls suggests). What is certainly more advanced is the learning component of the system: the more the user chooses a certain answer option for a certain caller, the higher in the list of options it will show up the next time. This means, if a user *always* answers a call from her mother in law with "I am busy, I will call back later," then this option will soon be on the very top of the list that is presented to the user when she calls.

An interesting variation of the above systems is IDEO's *Social Mobile* [86] #2, the speaking mobile. Instead of the user sending back digitized or even synthesized speech triggered by buttons or text menu entries, IDEO's concept suggests sending back to the caller simple but expressive sounds that are controlled with a kind of joystick. It seems like the user would intuitively imitate simple prosody patterns of a pseudo conversation.

All the above-mentioned systems do something very important and interesting: they allow a kind of multi modal interaction, or rather, modality crossover, by deferring a caller to a different medium. However, there are several important differences to an Intermediary:

- Even SoloMio's advanced Interactive Voice Response is not an independent entity by itself—there is no independent intermediate party that can act on behalf of the user. It is more like a 'audio puppet', completely controlled by the user
- The alerting problem is not addressed at all, and essentially is still the same as what we already have nowadays: even if a visual blinking alert is less disruptive, it is not different from ringing or vibration, and certainly not as sophisticated as human style non-verbal cues.
- Personalization described by SoloMio consists only of having lists (per caller ID) of responses and ordering them for highest frequency in use.
- None of the above systems have an idea of content of the call, a characteristic typical for passive context-aware systems.
- An Intermediary is not only an active context-aware system, it is also able to participate in two conversations simultaneously.

This capability is unique, and requires the Intermediary to be an autonomous entity.

6.3. Social intelligence in software and robotic agents

In the domain of Socially Intelligent Agents, the following works are relevant or related:

- Kihlstrom (2000) [98]
- Dautenhahn (1998, 1999, 1999b, 2000, 2000b) [34][35][36][38]
- Bickmore (1999, 2000, 2003) [13][15][14]
- Michaud (2000) [135]
- Lockerd (2002a, 2002b, 2003) [119][121][120]
- Gong (2002) [68]
- Hogg et al. (1997, 2001) [81][82]
- Duffy (2000, 2003) [46][47]
- Ball et al. (1997) [4]
- Berner et al. (2000) [11]
- Nicolescu et al. (2001) [148]
- Pynadath et al. (2000) [162]
- Scerri et al. (2000) [173]
- Scassellati (2000) [172]

Recognizing people

Wouldn't the Intermediary's embodiment have to be able to recognize at least its user? Probably it should recognize the user, or even the people around it:

> "It is then argued that social intelligence is not merely intelligence plus interaction but should allow for individual relationships to develop between agents. This means that, at least, agents must be able to distinguish, identify, model and address other agents, either individually or in groups." (*Modeling Socially Intelligent Agents* Edmonds, 1998) [53]

In the current implementation, the embodiment itself does not recognize people close by; however, the Intermediary may be given the identities of all conversational partners via the Conversation Finder nodes.

Recognizing the people with their respective relative location to the user is only possible via sophisticated face recognition (computer vision) or speaker recognition (audio).

6.4. Conversation finder

The Conversation Finder nodes give the Intermediary rich contextual information about the user's social state: if she is all by herself, or part of a group, mainly listening to a speaker, or being the main speaker herself, the size of the conversational group, as well as the ratio of listening to talking participants.

6.4.1. Alignment of speech

The following authors have researched the use of alignment of speech to determine conversational groups:

Egbert (1997) [55] describes *schisming*, the transformation of a conversation with four or more people into smaller, simultaneous conversational clusters, which is the underlying phenomenon used by Conversation Finder to detect conversational groups.

Aoki et al. (2003) [2] describe as system that clusters conversations in a multi-party audio chat. The main difference to Conversation Finder is that their system determines schisming centrally, where Conversation Finder employs a completely decentralized approach.

Choudhury et al. (2003a, 2003b) [28][29], Choudhury (2003) [27], Clarkson (2002) [30], Eagle et al. (2003b, 2003c) [50][51], and Basu (2002) [9] are all doing some kind of conversational analysis that relies on alignment of speech, but they do it offline and centralized. Conversation Finder is a decentralized real-time system.

Holmquist et al. (2001) [83] suggest the concept of *context proximity* that is also used in Conversation Finder. For finding conversations, it uses—in addition to alignment of speech—a second, less obvious and implicit criteria: being close enough to be able to have a conversation at all. This is implemented via RF transceivers that have only very limited range. If a person is out of range of the transceiver of a user, she—with all likelihood—is also too far to be a participant of the user's conversation.

Vertegaal et al. (2001) [203] describe work that uses vision to detect conversational groups instead of audio.

6.4.2. Sensor node networks

Conversation Finder relies on a set of wireless sensor nodes, or more precisely, a network of wireless devices with sensors. The following researchers (among others), have done related work or built related systems:

- Rhee et al. (2002, 2003) [165][166]: *i-Beans*, commercialized by Millennial Net (http://www.millennial.net).
- Poor (2001, 2002) [158][159] *EmberNet*, commercialized by Ember Co. (http://www.ember.com/)
- Kasten et al. (2001) [93] *Smart-Its*
- Berkeley's *Motes* (COTS, weC, Rene, Dot, Mica, Spec), described in, e.g., Hill (2003) [78]. Some of them commercialized by Crossbow (http://www.xbow.com/)
- Kahn et al. (1999) [91] *Smart Dust*
- Beutel et al. (2003) [12] *BTnodes*
- Lifton et al. (2002) [112] *Pushpin Computing*

- Gerasimov (2001, 2002) [64][65]: the *Hoarder board* influenced the design of the Conversation Finder node significantly, especially the transceiver section and the audio daughter board (see also DeVaul et al. 2003) [43].
- Kortuem et al. (2001) [100] describe a decentralized and self-organized network of autonomous, mobile devices that interact as peers

6.4.3. Social polling

The concept of social polling, or more specifically, the idea of a software agent or an appliance polling people in close proximity in a subtle way about the social appropriateness of interruptive events, seems not researched well at all. The main reason is probably that there is no common infrastructure for subtle polling. The current Intermediary system uses a vibrating finger ring as a *pre-alert*—wirelessly actuated—if an interruption from a mobile communication device is about to happen. All involved people are given the possibility to “veto” an incoming communication in an equally subtle way by touching their ring. Such a small device with low-range radio two-way communication and silent actuator has not been built up until now.

However, there are two cases that come close to the idea behind social polling with subtle pre-alerts (although not including the possibility to “veto”):

The experience of pre-alerts might be similar to carrying a small vibration alarm that registers any cellphone activity in the room. (These devices are intended to alert the user if her cellphone is ringing. Therefore, registering *all* cellphone activity is an unavoidable bug due to the fact that the receivers within these vibration alarms can't be calibrated to only respond to the user's cellphone, but rather get triggered by *any* radio signal within a specific spectrum in close proximity of the user.) These vibration alerts usually occur a few seconds before a phone starts to ring, which allows the owner of the device to "magically" be able do predict an interruption. Such alerts are subtle, but also private, and of course the users cannot influence the interruption in any way.

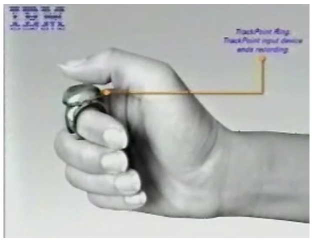


Figure 72: GlowRing

Also related to pre-alerts is the experience of a group of people having a phone conference on a speakerphone: each time, a few seconds before a cellphone within this group is about to ring, a few clicking noises are audible through the speakerphone, caused by radio interferences. These low volume clicking noises are unintended pre-alerts for immediately upcoming cellphone interruptions, but detectable only by participants who know about this effect. Although they are subtle and public, the participants still do not have any way to influence the upcoming interruption.

Miner and Miner et al. (2001, 2001) [139][140] have done some related research. In their Digital Jewelry project, they describe several versions of finger rings, both as input and output devices. Their LED *GlowRing* (Figure 72) glows upon an incoming email message in varying colors, depending on the importance of the message. When the user touches the face of the ring, it sends a wireless signal to the user's LCD bracelet to display the face identity of the sender, and another signal to the user's earring (which serves as a wireless headset) to play back the urgent message into the user's ear. The user can reply by using the microphone built into her necklace. It is not clear, though, how much of this scenario has already been implemented. Note also that the core idea behind social polling merely uses peripheral alerting similar to GlowRing, but eventually is concerned about a problem that the Digital Jewelry project does not address: the collective responsibility for interruptions from communication devices.

6.5. Issue detection

> "The concept of an intermediary that would act as an agent doing things you wanted done still thrives today. Still, I am dreaming of agents that can understand and interpret high-level goals and purposes. What is important? What is correct? What should be done? I want an agent to remind me, "hey boss but yesterday you said..." or "Professor, you want me to lie to the IRS..." or "But, honey that's wrong..." (Oliver Selfridge, http://www.almaden.ibm.com/almaden/npuc97/1997/selfridge.htm)

What Oliver Selfridge—who is credited by Licklider (Licklider et al., 1968) [108] to have invented this idea—describes here, is the ultimate Intermediary. This thesis work does not intend to realize such an omniscient agent, but parts of Selfridge's description come close to what the Issue detection module is trying to accomplish.

The Issue detection module is based on the idea that the calling party can provide important social intelligence clues about how an Intermediary should deal with an interruption for the user. Whereas the other modules of residual social intelligence try to harvest leftover social intelligence, the Issue detection module is using the content of the call directly to determine the relevance, timeliness, importance, etc of an interruption.

This is a very challenging problem, since it seems that the Intermediary has to understand the very content of a call. This sounds like an A.I. complete problem, especially since it might require natural language understanding from noisy speech recognition over phones. However, there might be ways to get content information out of call without solving the A.I. problem as a whole, e.g., by reducing the problem to the question if a given call is related to "issues" that are on the user's mind right now. This problem has two sides: first, determining what is on the user's mind (the issues), and second, analyzing the call to determine if it is related to some of these issues.

There are several domains where related research has been done:

Email and voicemail filtering
A possible success criterion for such a system could be if its filtering performance is better than current systems like CLUES filtering (Marx et al., 1996) [128] which look only at caller/sender ID, subject lines, threading (replies to replies), etc. Other research in the filtering domain, e.g., Horvitz et al. (1999) [84], tries to infer expected

"criticality" from incoming email messages. Both projects are related because they try to classify an incoming communication in terms of how relevant (or timely or critical) it is to the user—the same problem Issue detection would like to solve by looking at, or rather, listening to the content of a call.

Commonsense and fuzzy inference
Closely related to the Issue detection module is research that tackles the problem of how to use commonsense to do fuzzy inferences and query extension (e.g., Lieberman et al., 2003 [109], Liu et al. 2002 [115]), as well as topic detection (Eagle et al., 2003d) [52]. All these projects are based on the idea of using commonsense knowledge and commonsense reasoning (Minsky 2000, Minsky et al., 2002,) [142][141] to eventually improve the usability of computers. Projects that have collected commonsense via a Web interface are *Openmind* (Singh, 2002) [192] and *LifeNet* (Singh et al., 2003) [194]. The Issue detection module could rely on this huge repository in the form of English sentences to do fuzzy inferences. The module could alternatively use a more compact semantic network like *ConceptNet* (Liu et a., 2004) [117] that is mined from the Openmind repository. (Liu et al., 2004, use the same method for textual affect sensing—a feature that would be an interesting extension of the current Intermediary.)

Kim at al. (2003) [99] are working on generating user hierarchies of interests, work which is very related to the Issue detection module.

Artificial minds
In a wider sense, some works that are concerned with creating artificial minds are related, e.g., Clocksin (2000) [31] and Davis (2000) [40]. The former gives priority to social relationships as a key component of intelligent behavior—an idea very related to Issue detection.

Mindmaker[6] offers several commercial products that are very relevant to the Issue detection module. Their portfolio includes a pre-processor for Internet/Intranet keyword search engines, an automated document highlighter and summarizer, a text classification and categorization engine, and other. Although their products *FlexAnswer* (a Web-based question-answering system that learns from previous questions and answers) and *Tell A Phone* were supposed to help gather content from a conversation, it is not clear how well these products will live up to their promised performances (if they are still available, which is not clear from their web site).

6.6. Animatronics

6.6.1. Physically embodied agents, RUIs

An important aspect of this thesis work is the embodiment of the Intermediary, and there is a wealth of related work in the domain of physically embodied agents:

[6] http://www.mindmaker.com/

Robotic User Interfaces

A *Robotic User Interface* (RUI) (e.g., Bartneck et al. (2001a) [7], Bartneck et al. (2001b) [8], and Sekiguchi et al. (2001) [187]) is the paradigm where robots are used as an interface between the physical world and information world.

Kuzuoka et al. (1999) [102] and Greenberg et al. (2000) [74] suggest digital but physical surrogates in an office environment (Figure 73). They are digital representations of people (avatars), something the Intermediary does not intend to be. Jabarin et al. (2003) [90] suggest the *eyePHONE*, a mechanism to initiate and respond to communication via eye contact. Although it is also based on the avatar paradigm, it uses the strong social cues of eye contact, a feature that this work shares.

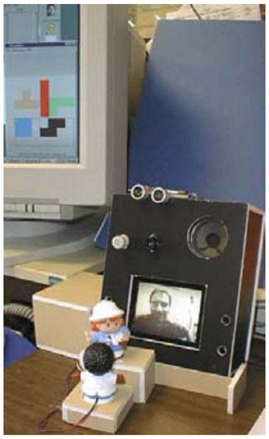


Figure 73: Digital but physical surrogates

Embodied Agents with Social Intelligence

This work is also in the tradition of *Embodied Agents with Social Intelligence*, which may be based on Reeves at al. (1996) [164] seminal findings about how humans treat machines as social beings ("computers as social actors") [20].

Kismet (e.g., Breazeal, 2002 [19], Breazeal et al. 1999 [22]) is a prominent example of a socially intelligent robot. Although it has not the same function as an Intermediary, it demonstrates the importance of socially strong nonverbal cues to grab attention, show interest, etc. Breazeal et al. (2000) [21] found that "humanlike eye movements of a robot have high communicative value to the humans that interact with it. This can be a powerful resource for facilitating natural interactions between robot and human" since humans seem to be hardwired to react to facial stimuli, and a socially intelligent robot should take advantage of that. Kismet also shows how fluid the non-verbal interaction between human and embodied agent is. Human conversation is not like playing chess with discrete turns, but it is like a dance—a fluid, and continual regulation of the other's non-verbal and verbal behavior. Kismet excelled at the non-verbal "dance," and that's why interacting it is so compelling.

Although Duffy (2000) [46] distinguishes between *social* and *societal* robotics—robots being social with each other, versus the social integration of robots into human society—and focuses in his thesis on the former, he later acknowledges the importance of "meaningful social interaction between robots and people through employing degrees of anthropomorphism in a robot's physical design and behavior." (Duffy 2003) [47]

Okada et al. (1999) [151] look at the important social bonding between artificial autonomous creatures (such as cyber pets) and humans, especially its conversational aspects. Fong et al. (2002, 2003) [57][58] present a very thorough survey in the domain of socially interactive robots.

Green (2001) [72] studies natural language interfaces for domestic devices, and suggests characters embodied as interface robots that are collocated with the appliances they interface with, in order to make talking to, e.g., the VCR or the microwave oven, more socially acceptable. The same author is also working on a related project (Severinson-Eklundh et al. 2003) [189], where a small, embodied agent is used as a representative for a bigger mobile robot (Figure 74), serving as a natural communication partner. The embodied agent can also be interpreted as the "driver" of the robot. Although the mobile robot itself is not humanoid, the small figure has some simple human traits, with a head, arms and a body.

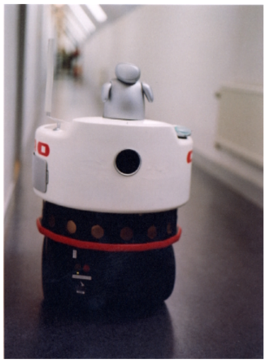


Figure 74: Cero robot

Subtle Expressivity for Characters and Robots

Suzuki et al. (2003) [196] initiated work under the label of *Subtle Expressivity for Characters and Robots.*

This idea seems to resonate significantly with Hansson et al. (2001) [76] work on subtle but public alerts in communication. Two relevant papers in this context are Liu et al. (2003) [118] and Isbister (2003) [89].

Intelligent Interface Agents

This system is also related to *Intelligent Interface Agent* research. The first embodiment of a complex Intermediary as a bird was probably the COMRIS parrot (Co-Habited Mixed Reality Information Spaces, Figure 75), a somewhat related project that was conducted at Starlab (e.g., Van de Velde 1997, De Haan 1999) [199][41]. It is a wearable advisor, attempting to create moments of interest for its wearer, in the context of a large-scale event (conference, fair). It delivers a series of spoken messages (signals) by which it influences its wearer's behavior. Van de Velde (1999) [200] looks at the effectiveness of such a wearable as an advice giver. Geldof (1999) [62] and Geldof et al. (2001) [63] look at how natural language of such a wearable could decrease its intrusiveness.

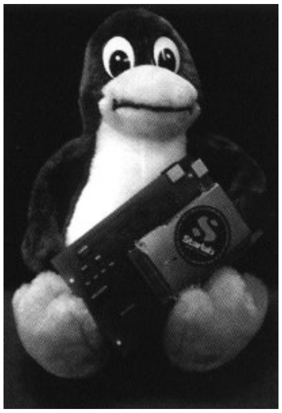


Figure 75: COMRIS

Kaminsky et al. (1999) [92] describe software tools for Programmable Embodied Agents (PEA) which are "portable, wireless, interactive devices embodying specific, differentiable, interactive characteristics. They take the form of identifiable characters that reside in the physical world and interact directly with users. They can act as an out-of-band communication channel between users, as proxies for system components or other users, or in a variety of other roles." This work is very related, since the authors use robotic character toys as a hardware platform for their software widgets, and use this system both as a channel and a proxy of a person, device, or event.

The success of our embodiments may be related to being cute stuffed animals, which makes them rather distinct from the stereotypical cold robot. In related work, Yonezawa et al. (2001) [167] describe a sensor doll for musical expression that is capable of multi-modal interaction with the user. The doll is a simple embodied agent with a mind of its own that displays its own built-in autonomous behaviors when responding to external stimuli. Although this work uses a physical embodiment for the agent, the output of the system is rather music and audio than physical movements.

Also a doll, a teddy bear, is used in *RobotPHONE* (Sekiguchi et al., 2001) [187], which seems to solve a similar problem as the Intermediary's animatronics, but follows an orthogonal approach: the caller manipulates directly her doll, and this manipulation is transmitted unmodified to the user's doll, and vice versa. This means, there is no agency that mediates between caller and user, which is an essential element of an Intermediary system. The Intermediary's animatronics are entities independent from caller and user, whereas Robot-PHONE does not make that claim.

6.6.2. Gestural interfaces

This work looks at how human gestures can be used in user interfaces. Most often, it is about the inverse of an animated figure that uses gestures:

> Making the Physical Environment Interactive: Tiny motors and sensors will make physical objects interactive and create a renaissance for gestural user interfaces. As interface design moves from the screen to the material world, the need for simple, easy to use designs will only increase."
>
> "(…) Of course, the computer could also express its side of the dialogue physically. Presenting facial expressions on a moving doll is a much more promising user interface component than simply pasting the expressions onto a GUI, as in projects like Boo and Ananova."
> (Jakob Nielsen's Alertbox, August 5, 2002, http://www.useit.com/alertbox/20020805.html)

Arguably Microsoft©'s most innovative product of the 1990s was ActiMates *Interactive Barney* [195] (Figure 76), a plush toy containing a computer. When Barney's toe is squeezed, it sings a song; when his eyes are covered, it plays peek-a-boo. Barney is a combination of products that work alone, with a PC, or with the TV (with specially encoded VHS tapes or the television show "Barney & Friends" on PBS). The system focuses on learning readiness and early learning skills to make learning fun and fascinating for preschoolers.

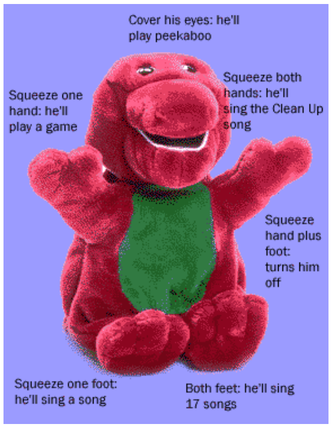

Cover his eyes: he'll play peekaboo
Squeeze one hand: he'll play a game
Squeeze both hands: he'll sing the Clean Up song
Squeeze hand plus foot: turns him off
Squeeze one foot: he'll sing a song
Both feet: he'll sing 17 songs

Figure 76: Interactive Barney

Barney is a 13-inch animated plush doll. Arms and head are actuated, and it has audio output. It has a set of five sensors: 4 touch sensors (one in each hand and foot), and a light sensor located in his left eye. It can react to user input by moving and speaking pre-recorded, digitized speech and programmed motion. Barney can be interfaced via internal RF transceiver. When the counterpart transceiver is attached to a TV and VCR, he can receive new speech and motion from encoded videotapes that play as the child is watching the video. If the transceiver is attached to a PC, the data link is duplex, and it can both receive new speech and motion content from the computer and transmit sensory input back to the computer.

Buddy Bugs (Figure 77 left) [73] is an ambient peripheral physical interface that represents Windows® Instant Messenger contact list, where people are represented by glass bugs on a leaf. A bug lights up and moves about depending on the status of the person it represents.

Ele-Phidget (Figure 77 right) is an ambient notification for an audio chat program. When one receives a message, the elephant turns around and faces the user. The user pushes the elephant's stomach to listen to the message. When no messages are left, the elephant turns away. To record a message, one squeezes the elephant's head and speaks into the elephant's trunk. A second squeeze stops recording and sends the message.

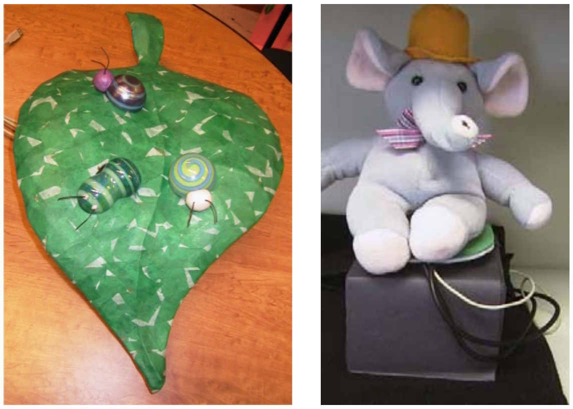


Figure 77: Buddy Bugs (left), Ele-Phidget (right)

6.6.3. Dolls and toys

During the last few years, a variety of toys were commercialized that share one or several characteristics with the Intermediary embodiment described in this thesis work.

Cellphone covers
Although no animatronics, cellphone covers such as *FunFriends*™ (Figure 78) and *Cellbabies*™ transform a bland cellphone into a cute fuzzy animal. The intended effect may be that that the cellphone should evoke zoomorphic or anthropomorphic reactions, and therefore co-located people may be less annoyed by the interruption:

> “The 1st Cellbaby™ was born on August 4th, 1999 on a common stroll down 5th Avenue in New York City. Everywhere I went I could see how irate people became as soon as someone pulled out a cellphone. What was it, I wondered? Something was wrong. Why would a communication device offend people and make them so angry and upset. Something was missing. That's when a thought came to mind and psyche gave me a cute little Cellbaby™. I didn't realize that it would have such an amazing impact. Everywhere I went whether it was with Betty the Bunny(tm) or Dylan the Dog(tm) it seemed as if everyone responded with joy when I pulled out my Cellbaby™ cellphone. No longer was I scorned in the coffee shop or abused on the bus, everyone loved my Cellbaby™. "Oh it's soooo cute!" people would say with a big smile on their face. That made me so happy.” (http://www.cellbaby.com)

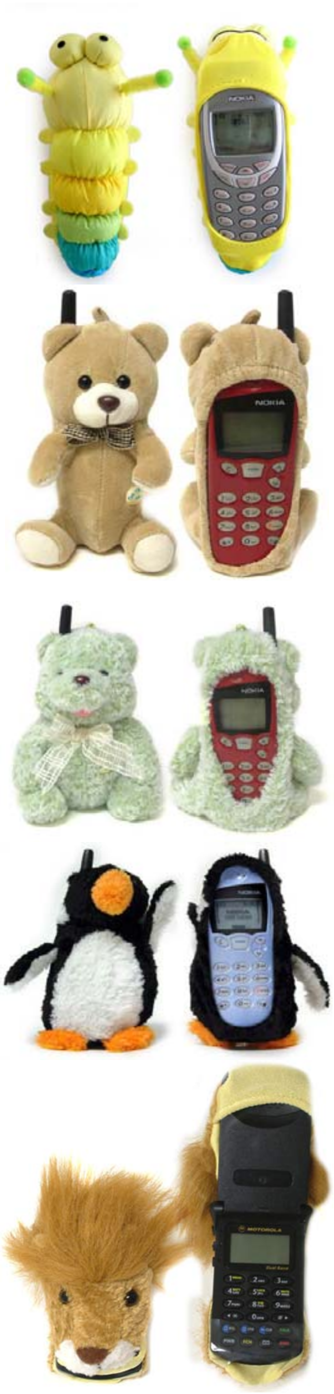


Figure 78: Animal cellphone covers

Cellphone call summarizer
Specifically in the Asian markets, there are products that may be based on the same idea, but go beyond cellphone covers. *MobbyPet* (Figure 79), a small Teddy Bear figurine that is connected to a Japanese

cellphone, which summarizes to the user in spoken language who has called and how often. This appears to be the first “agent” like teddy that specializes on mobile communication, and even uses spoken language for a specific purpose. Nevertheless, this system is not real-time, it merely summarizes past events.

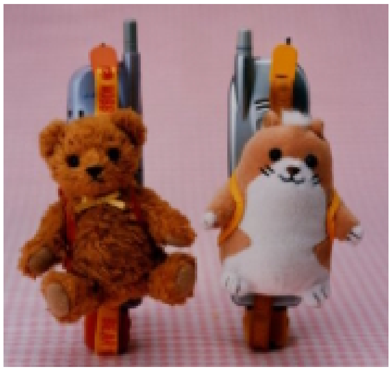


Figure 79: MobbyPet

Cellphone doll as phone attachment
So-called ”character phones” (Figure 80) with hands free voice dialing allow a user to dial a number by saying the callee’s name. Upon an incoming call, the animated handset lowers its arm to give the user the handset. The animation synchronized with incoming and outgoing calls.

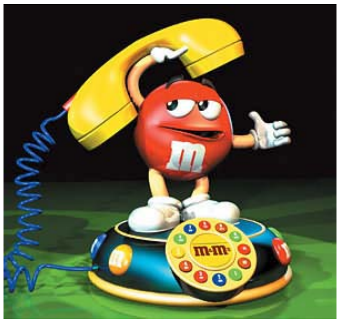


Figure 80: M&M™ One Character Phone

Why are nonverbal cues are better than just using digitized voice, like with voice chips (cheap, ubiquitous speech synthesis chips integrated into products to give them autonomous voice)? In other words: How is the embodiment of an Intermediary different from a character phone with voice dialing and speech recognition?

Experience shows that talking toys are perceived as annoying and/or silly after a relatively short time. It appears that the idea of a “talking” avatar (both physical and software) may be problematic in real life. The Intermediary does not rely on speech, but on non-verbal communication and cues, because they are intuitive and less intrusive.

Cellphone call alert holders
These desktop dolls serve as a base station (and sometimes charger) for a cellphone, and alert the user to an incoming call while the phone sits in the doll's pocket. The animated doll (Figure 81) can dance to a song (ring tone), but stops once the user pulls the cellphone from the pocket.

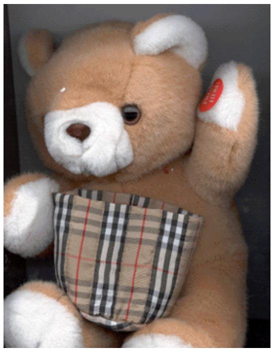


Figure 81: Cellphone call alert holder

As with cellphone dolls, the audio is often perceived as annoying and silly, maybe because speaking avatars are disappointing. But if the dolls' behavior would be subtler, and had nonverbal capabilities beyond just invariably waving, it might be more acceptable.

Robotic dolls connected to desktop computer
Interesting “cousins” of the Intermediary embodiment are the *Dot Pals (E-Pals)*, animatronic characters delivering wacky actions and commentaries in reaction to tasks performed on computers.

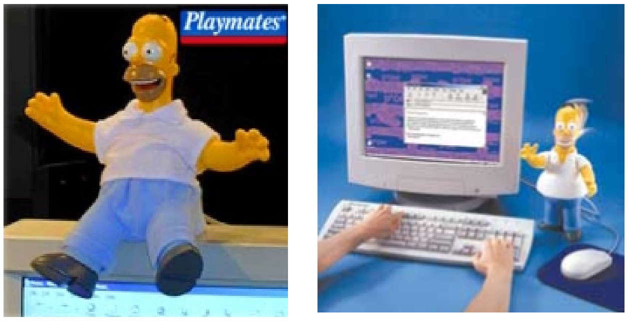


Figure 82: Dot Pal

> "This interactive figure, standing about 9 inches tall, hooks up to your computer and openly shares his thoughts about what you're up to. It recognizes most common desktop applications and will chime in, for instance, with his thoughts about your decision to print a document. Beyond merely talking, he also features some limited motion. Dot Pals know the date and time, and can remember personal information users have fed into the desktop calendar, including special occasions like birthdays or anniversaries. As you work on your computer, commands trigger appropriate responses as frequently as you choose. For instance, a Homer Simpson Dot Pal might respond to a spelling error by saying, "D'oh!" or get excited when he sees mail in your in-box with a familiar "Woo-hoo"!
>
> "Interactivity control: You can globally control the level of interactivity for your Dot Pal and the supported applications by dragging the control to the desired level from "Rarely" to "Always."
>
> "Spontaneity control: This control will select how often you would like your Dot Pal to randomly say something while your computer is running regardless of whether there is any input from the keyboard or mouse. You may also turn this feature off by clicking on the "Off" box to put a check there." (http://www.hedges.org/Simpsons/toys.html and others)

The Dot Pals have only limited gesture capabilities. Although they move, their non-verbal communication is not supposed to be unobtrusive via. There is no advanced agency in the backend that would justify their behavior/interruptions. They may be fun for a short time, but then become annoying. The randomness of the interruptions could be interpreted as 'life-like,' but may be perceived as socially inappropriate.

Another cousin is *PC Mascot*[7], Figure 83). It seems to come even closer to the idea of the *Familiar* on the user's shoulder, since it is indeed a talking parrot. Nevertheless, it is on the same level as the Dot Pals, since the animation is not used to unobtrusively get attention.

[7] http://www.pc-mascot.com/

The PC Mascot lets the user know via movements when new email messages have arrived, retrieve them, and even read them to the user. It also reminds the user of scheduled events and anniversaries. It can also tell jokes, tongue twisters and makes other 'witty' remarks.

It also has a random phrase mode, which is probably the most annoying thing about this product, as the following user reports show:

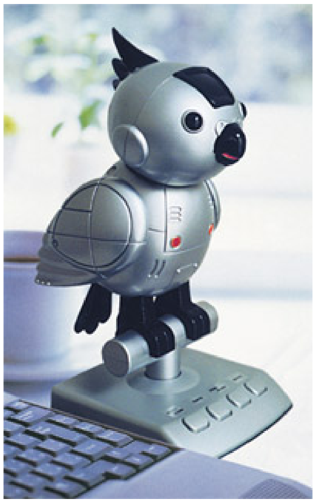


Figure 83: PC Mascot

> "'Special Agent PC Mascot' is it's name, annoying the pants off you is it's game. Mitsumi have tried something a little original, and a little out of the ordinary as far as their usual products are concerned.
>
> (…) It sounds like a fun novelty item, but after a while the novelty wears off and you'll be ready to throttle the poor bird. (…) Well I've referred to this annoying voice, but you're yet to hear it, so I've got a little treat for you. There are in-fact five voices to choose from, but from what I can tell they're simply the same voice at different tones. You can use the software to change the volume and speed of the speech, which is handy. 0% volume is my favorite value. 'A PC Mascot is for life - not just for Christmas!' 'It's so boring to sit here all day long! Come on, please play with me.' 'I'm so lonely, can we be friends?'
>
> "Now imagine that every other minute, or alternatively play them over and over again. I think you should begin to see how irritating the PC Mascot is. But wait! It gets more annoying. On top of the voice, you have the blinking lights, flapping wings, wagging tail, tilting head and moving mouth. All of which move very noisily and very jerkily.
>
> "(…) The PC Mascot is a novelty item. It starts off as a bit if fun, but it then gets very annoying the more time you spend with it. It's made of plastic, and when it moves makes loud mechanical noises. The synchronization between the voice and the movement of the mouth is surprisingly good. If Mitsumi were trying to design a cute animal, I would have advised them to make it fluffy, make the movement gentler and make the voice far nicer. In my opinion paying £40 for a lump of plastic that does the job of free software, and costs you hundreds of pounds in anger management therapy seems a bit silly. Hats off to Mitsumi for trying to develop an original and novelty item, but it would certainly need some improvements before I'd put it on my shopping list."
> (http://web.archive.org/web/20030602171543/http://www.blagged-hardware.net/index.php?index=32)

Phone dolls

Although not animated, *Wabi* (which stands for While Away, Be In-Touch) (Figure 84) is the unique combination of a plush toy (three bears, a teddy, a polar, and a panda) with an integrated one-way cordless phone. It allows parents to leave messages for their child. Once a message arrives to the bear, the toy makes a giggling sound. The child can then retrieve the message by pressing the bear's right paw. Parents also can purchase stories for the bear by paying for them ahead of time through Wabi Inc., which controls the backend infrastructure, including the delivery of all messages. To listen to the stories, the child would press the bear's left paw.

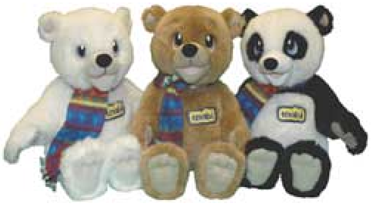


Figure 84: WABI

7. Summary and Outlook

This thesis work is the apex of a thread of research that I started exactly twenty years ago. It began with observational psychological studies, and culminated in the implementation and evaluation of a radically novel concept for telecommunication, an Autonomous Interactive Intermediary. I will summarize this path in section 7.1.

Although well researched, my implementation of an Intermediary is still on a prototypical stage, and it would be interesting to carve out how Intermediaries—or rather, some features of them—could find their way into mass products (section 7.2).

There are many ideas that were left out both when I developed the underlying concept and in my implementation of an Autonomous Interactive Intermediary, for various reasons. I will describe some of these longer-term future works in section 7.3.

7.1. Research summary

The main domain of this research is human computer interaction; within this domain, the focus is specifically on the following areas:

- Computer-mediated communication (CMC), specifically mobile communication
- Agents and AI, specifically agents with commonsense and social intelligence
- Ubiquitous computing and sensor networks

As a researcher, I am interested in these areas from the user and user interface perspective: What kind of mobile communication devices serve people the best? What kinds of user interfaces are socially acceptable? How would people interact with socially intelligent agents? How would people react to entities with real commonsense? How can we build agents that increase their social intelligence and commonsense reasoning capabilities during interactions with people? How does ubiquitous computing and personal sensor networks help mobile communication and user interface agents?

In addition to developing a solid theoretical basis for research, I am convinced that it is important to build prototypes of systems and design appropriate user interfaces to see how real people interact with and react to such new technologies. Prototypes need to work well enough in the real world so that we can evaluate how the new paradigms affect people's lives.

7.1.1. Psychological impact of telecommunication

In the mid eighties and early nineties, I studied how people use telecommunication technologies. My empirical and case studies focused on the psychological aspects ad impacts of telecommunication media choice.

In my final two-year study which earned me a Master's degree in Psychology, Philosophy and Computer Science, I examined both the communicative behavior in general and the use of communication technologies of eight subjects in detail using extensive problem-centered interviews. From the interview summaries, a general criterion for media separation was extracted, which allows the systematic separation of all media into two groups: on one side the verbal-vocal, realtime-interactive, and non-time-buffered media like telephone, intercom, and face-to-face communication; on the other side the text-based, asynchronous, and time-buffering media like letter, telefax, and email. The two media answering machine and online chatting (realtime communication via computer monitor and keyboard) occupy exceptional positions because they cannot be assigned to either group. Therefore, these two media were examined in detail. Moreover, through analyzing them under the aspects of both a semiotic ecological approach and a privacy regulation model, important characteristics and phenomena of their use can be explained.

7.1.2. Agents for mobile communication

In the past eight years, my interests shifted from psychological studies to engineering, and I started building agent systems for mobile communication. My master's thesis at the MIT Media Lab was a software agent that deals with incoming text messages, and forwards them in an intelligent way to the user's mobile devices. It can send a message to several channels in turn (Figure 85), waiting in between and monitoring the progress of each message, the final goal being the message is delivered in the most efficient way. The higher the importance and timeliness of the message, the more 'expensive' and persistent communication channels the agent is allowed to use. For example, if the user does not read an extremely important message on her cellphone, pager, and computer after some amount of time, the agent can send the user a fax to her current location, or even call her up on landline or mobile phones to deliver the message. If the message is not important, the agent does not bother the user at all, since she will read the message when she checks email on her computer.

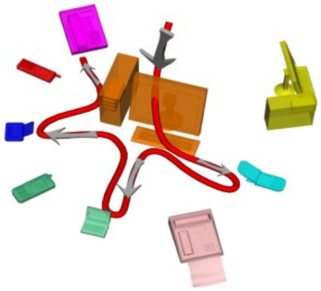


Figure 85: Message routing in Active Messenger

7.1.3. Agents with social intelligence

During the last three years, I extended the focus of my work on mobile communication agents to synchronous voice communication. Given increased wireless bandwidth, mobile communication's main problem is now that not every incoming communication attempt—be it text or voice—may be worth interrupting the user. Such interruptions will disrupt our ongoing face-to-face interactions. Currently, our mobile devices do not 'care' about our ongoing conversations, about the relationship between the caller and the user, or even what the purpose of the call is. In order to be more socially appropriate, our mobile devices need social intelligence:

Intermediaries and dual conversational agents

Since our mobile devices become more pro-active and autonomous in their behavior, they may become *Intermediaries* that can stand between the user and the people trying to contact the user. An Intermediary deals with incoming communication attempts when the user cannot or does not want to. It's a *dual conversational agent* that can converse with both user and caller simultaneously, mediating between them.

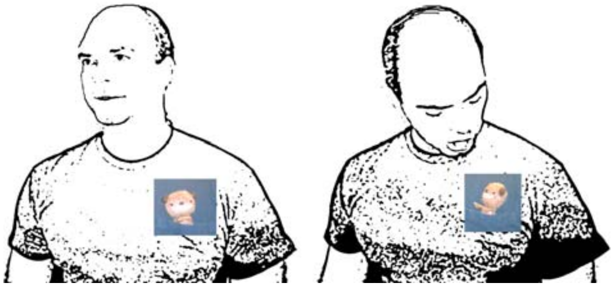


Figure 86: Mockup creature in chest pocket

Embodiment of Intermediary

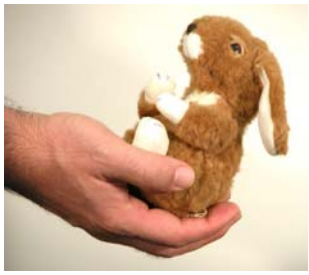


Figure 87: Intermediary embodiment

Current mobile communication devices do not grab our attention in a socially appropriate way—they could be disrespectful of ongoing social activity such as an important meeting or private dinner. To improve on this, I have built a system where the Intermediary is embodied in a small portable animatronic device (Figure 87), as a personal 'companion' for the user (Figure 86). This embodiment of the Intermediary is able to use the same subtle but still public non-verbal cues to get our attention and interrupt us like humans would do (like eye gaze and small gestures). The user can whisper and listen to her Intermediary, receiving and replying to voice instant messages. If the user wishes, she can also bypass the Intermediary altogether and get into a synchronous voice communication with the caller via talking to the embodiment.

Harvesting 'residual social intelligence'

Making truly socially intelligent devices is no easy task. In addition to behaving socially appropriately by using human style non-verbal cues, I have developed a system where an Intermediary harvests 'residual social intelligence' from human and other sources (Figure 88). These sources include:

1. caller (remote person)
2. callee (owner of mobile device)
3. co-located people
4. owner’s current location

E.g., the Intermediary can poll co-located people unobtrusively if an interruption at the current time would be appropriate. For that purpose, all people involved in a face-to-face conversation with the user can anonymously veto to an upcoming interruption—without knowing whose communication device is about to interrupt. Another source of social intelligence is 'room memory,' a sub-agent that remembers whether or not occupants usually pick up calls at this specific location at a given time.

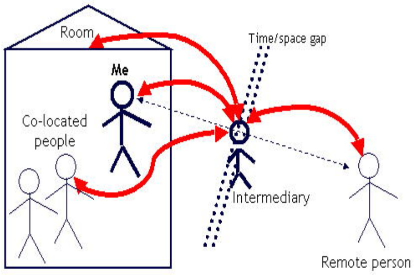


Figure 88: Sources for social intelligence

Sensor network

I have implemented these ideas for harvesting social intelligence using a decentralized network of custom wireless sensor nodes (Figure 89).

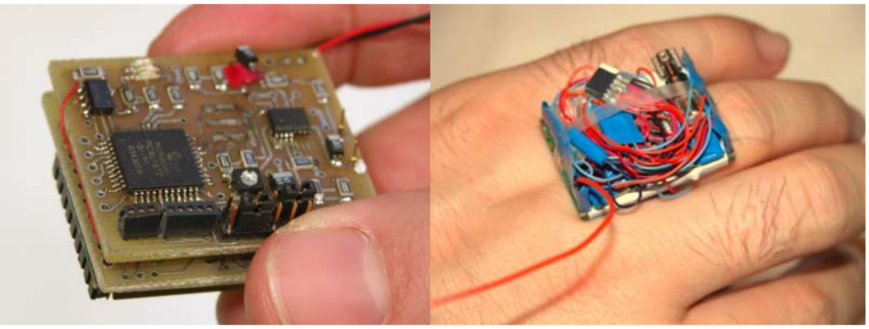


Figure 89: Conversation finder node (left) and wireless ring (right)

One type of node, the *Conversation Finder* node, is worn close to the neck of the user. It communicates with other nodes close by and in real-time determines the conversational status of the user, such as how many people are in her conversation, and if the user is mainly speaking or just listening. A second type of node is worn by the user as a ring. It alerts all participants in the same conversation of an incoming interruption with a slight vibration, and anybody can veto anonymously by touching his/her ring. This veto, as well as the current conversational status, is transmitted to the Intermediary and taken into account when trying to assess the appropriateness of an interruption.

7.2. Migration paths to mass products

The prototype technologies described in this thesis are research tools built for testing novel protocols and user interface paradigms, and are built in a modular and open way that allow debugging and extending the hardware and software in the most convenient way.

However, they are not on the same level as a commercial product in terms of reliability, range, battery life, and miniaturization.

There are many hurdles for how to port these prototype technologies to commercial mass products: some hurdles are of technological nature, some economical, some social, some political, and some legal. In the following section, I will focus on the first and second one.

On a very simplified (almost caricature) level, the technology of an Autonomous Interactive Intermediary can be described as follows:

"Take a cellphone, breed it with a cute animatronic animal, and add a bit of agent intelligence to make it behave right…"

Although grossly simplifying, this phrase shows two possible ways how to go about introducing an Autonomous Interactive Intermediary as a mainstream commercial product.

7.2.1. I/O device for cellphones

The first option is to build upon the aspect of its mobile phone functionality. For example, the Intermediary embodiment could be marketed as a Bluetooth extension to existing cellphones. The user can keep the cellphone in her backpack or pockets, and uses the Intermediary embodiment to take and make phone calls, etc.

My most recent embodiment, the squirrel, is already fully Bluetooth compliant, and it is conceivable that in the near future all required processes for an Intermediary could be run on a powerful cellphone with built-in Bluetooth. Although there are no such phones yet, I have no doubt that the mobile phone industry will eventually develop powerful enough hardware platforms that can run the conversational agent (including speech recognition) and the animatronics control software (maybe excluding the behavior creation tools), as well as a provide a wireless network connection to services that cannot run on the phone itself, or need access to other servers.

In its simplest implementation, such a device would be similar to a Bluetooth headset, except that it has built-in speakerphone capabilities, and alerts with movements instead of ringing and vibration. Obviously lots of mechanical and electrical aspects have to be addressed, such as how to make the actuators both reliable and quiet, and what size of power source is necessary to give it a battery life comparable to a cellphone.

Mechanically, it has to be sturdy enough so that dropping the animatronics from a table does not damage it. Similarly, the actuators need clutches that protect them from external manipulation. The currently used servos are not meant to be manipulated externally because of their internal gear trains. Either a mechanical decoupling is necessary, or other types of actuators should be considered, like the completely silent memory shape alloys, or voice coil actuators (McBean et al. 2004) [130].

7.2.2. Animatronic toy with cellphone functionality

Another migration path for commercializing an Intermediary could be to place it as a product in the toy market. There is an incredible variety of animatronic toys already available, and one could imagine extending the functionality of such a toy with some of the key features of an Autonomous Interactive Intermediary. There are already products that combine stuffed animals with cellphones or other wireless networking capabilities: these are possible first steps towards a more complex device similar to an Autonomous Interactive Intermediary.

Advantage of this path would be that the consumers (meaning, children) will have much less problems in adjusting to the switch from a cellphone as a passive tool to a proactive Intermediary, one of the social adjustments that may be necessary to allow an Intermediary to become a successful mass product.

7.2.3. Embedded sensor networks

However, both approaches described above do not take into account that a fully developed Intermediary relies on external hardware such as technologies that allow it to find conversational groupings (Conversation Finder nodes) and can poll bystanders with pre-alerts (Finger Ring).

Social polling, which is done via actuated wireless finger ring in this thesis, simply requires miniaturization of transceiver and actuators—there is no way around that. However, there are already products that take a first step into this direction, such as a finger ring that flashes when a call comes in. Such products are not able to distinguish between the user's phone and any other phone in close proximity, and serves as a rather primitive alerting device similar to the external pocket-sized vibration alarms that were made commercial a few years ago.

There may be other implementations of the idea of social polling, which are equally subtle and allow anonymous feedback. One option would be to build the social polling feature into the sole of a shoe. The wearer could simply touch the back of on of his shoes with his other shoe in order trigger a veto, and would receive pre-alerts via a vibration motor that can comfortably be placed in a sole.

The conversation finder infrastructure requires more sophisticated technology, though. Since it requires a microphone close to the user's mouth, it may be possible, however, to add the feature of finding conversational groupings into wireless headsets for cellphones. Such headsets, these days often on a Bluetooth basis, are already optimized to pick up the user's voice and canceling out ambient noise.

Such a headset would use the same underlying principles as my Conversation Finder nodes to detect how many people are on the user's conversation. It would listen for and send short messages when the user is talking, either via Bluetooth, or via a dedicated and much less complex (and cheaper) transceiver technology.

In general, the current sensor network architecture degrades gracefully when not all participants have are parts of the described infrastructure available.

7.3. Intermediary additions and extensions

In this section I will describe some possible extensions for an Autonomous Interactive Intermediary.

7.3.1. Direct user feedback and learning

One possible extension could be to add functionality to the dual conversational agent that allows the owner of a cellphone to give *direct feedback* to the Intermediary. It could include general policies such as:

> "Don't interrupt me in the next 10 minutes"

or more specific instructions such as:

> "If my father calls, connect him directly even if I am busy."

This could be done either from a textual interface (e.g., via a Web interface), or by talking directly to the Intermediary embodiment. The former would be easy to parse, if given a text form of some kind; the latter requires simple natural language understanding. However, since the domain for such interactions and instructions is rather narrow, a model of the user's spoken language input may be constructed to detect the owner's wishes properly and reliably.

This feature would allow the Intermediary to profit from direct instructions, which may increase its immediate usability. Obviously, if there were no such direct instructions or policies from its owner, the Intermediary would fall back on the sources of social intelligence as described in this thesis.

Another useful feature would be to make the Intermediary 'learn' from its user. The idea is that the owner has the possibility to give direct feedback to the Intermediary about appropriateness of past actions, by simply saying:

> "Don't do *that* again!"

This command would refer to the last action the Intermediary took, be it either in the non-verbal domain, or on the spoken language level.

Although such a learning feature would be very convenient for the user, it is more complex than the earlier described one where the user gives direct instructions to the Intermediary. That's because the Intermediary has to interpret the user's approval or disapproval, and make an intelligent guess about which actions it concerned.

Finding the right focus may be non-trivial, e.g., if the Intermediary is waking up, and the user would tell the Intermediary to stop what it is doing right now, the system first has to determine if the user means, e.g., the specific behaviors that were chosen to alert the user, or if it is about the current time that is inappropriate for interruption, etc.

Of course the problem space could be limited to one or a few useful dimensions, such as associations between caller identity and urgency of interruptions. For example, learning could be used to teach the Intermediary about the importance level of certain callers, or rather, preset the importance of certain callers to a permanent level, like the owner's family members will *always* get connected directly, or an annoying caller that earlier tried to trick the Intermediary into inappropriate behavior will *always* get sent to voicemail.

7.3.2. Clarify interruptions; important vs. time critical

Another feature to consider would be allowing bystanders to understand interruptions by clarifying them. The current system is optimized to avoid inappropriate interruptions. However, it may be interesting to see if the overall acceptance increases if the Intermediary tried to explain the Interruption to bystanders.

For example, if the owner of an Intermediary gets a urgent phone call from a hospital, the Intermediary may choose to interrupt even in a highly inappropriate setting such as a lecture, but at the same time tries to convey the importance to the co-located people.

How such clarification could be done in an intuitive way is not clear, though. As an ultimate solution one could consider 'standardizing' certain non-verbal signals and loading them with commonly known meaning. Such as, if the Intermediary behaves in a very peculiar way, all bystanders would understand the importance level of the interruption.

Yet another feature to add would be the distinction between important and time critical interruptions, similarly to Marx et al. (1996) [128]: some interruptions may be important, but it is not necessary to wake up the owner during the night. For example, if a foreign citizen wins the green card (work permit) lottery, it is a very important event, but certainly not timely. In contrast, if there is an announcement for leftover food in a common work area, such information is timely (the food might be gone in 10 minutes), but certainly not important enough to interrupt an ongoing meeting between employer and employee.

A related feature would be for the Intermediary to determine the relevance of call to an ongoing conversation. For example, if the owner of an Intermediary is in an intense face-to-face discussion with two colleagues about how to shorten a conference paper submission to the right size, and a common friend calls to tell her that he has found more information on the web and that it is *not* necessary to reduce the current draft, this information is important to the ongoing conversation, and therefore the call should be allowed to interrupt immediately.

In order to make such a feature work, however, the Intermediary would require more in-depth natural language understanding on both caller and callee side, which is non-trivial, especially in a multi-party conversation.

7.3.3. Commonsense reasoning; affect sensing

A more radical addition to the current Intermediary would be to add more universal *commonsense reasoning capabilities* to cellphones. Such capabilities would complement the current harvesting of social intelligence. One way of implementing such a feature would be to use sources such as the *ConceptNet* semantic network (Liu et al., 2004) [117] for all actions and interactions involving human users. There is a variety of ways in which this could be implemented.

Yet another important feature would be to make the Intermediary aware of the emotional status of all involved. What if it would know about the excitement of the caller, the frustration of the co-located people, the user's own current mood? Such affect detection is non-trivial, but could extend an Intermediary's acceptance immensely.

7.3.4. Migrating the Intermediary concept

On a larger scale, a future extension of this thesis work could be to 'migrate' the concept of Intermediaries that mediate between humans and other entities, to entirely different domains.

In this thesis work I developed the concept of a telecommunication Intermediary that mediates between a user, remote communication party, and co-located people, capable of multiple concurrent conversations. Once users are comfortable with the basic concept of an Intermediary, the concept may be lifter to other domain, and used to mediate between other entities as well.

One obvious option for extension would be if local and remote parties do not speak the same language. Since an Intermediary can downgrade a synchronous communication to a semi- or asynchronous one (passing voice instant messages), the additional delay of language translation might be acceptable. This idea could be called an *Interlanguage Intermediary*.

Taking the idea further, the remote party does not have to be human at all! I envision conversational and embodied interfaces for personal assistants that are built to interact with any type of complex technology.

For example, there could be an Intermediary that mediates between the user and her advanced home. The idea is that the embodied Intermediary serves as an intelligent link between the user and the house, the user either being within the house, or somewhere else (work, vacation)—the user's location seems actually irrelevant.

Idea is that the user may not want to know all the details about the status of her house, and would prefer just getting notified of certain important events. The Intermediary would abstract all low-level sensor information, such as window and door switches, status of all electrical appliances, output of surveillance cameras and microphones, temperature and chemical sensors, etc., and interact with the user on a higher abstraction level, using human-style non-verbal cues.

Obviously, depending on the user's preferences and a given situation, she still may want to have access to the low-level data (e.g., asking the question "Is the stove turned off?"), so the Intermediary would have to be implemented with adjustable autonomy to allow dynamic switching between different levels of shared control.

Other complex technologies would also be suitable for an Intermediary. In addition to a house, I could imagine embodied Intermediaries to cars, airplanes, and even spacecraft.

The latter example is especially interesting. A spacecraft is an extremely intricate structure that incorporates technologies that are used in houses, cars, airplanes, and more, which makes it exceptionally complex. It is the perfect scenario for a powerful Intermediary that can mediate between the many complex subsystems—most of them are autonomous—and the human inhabitants.

Science fiction has taken up this idea many times already. For example, in Stanley Kubrick's cult movie *2001: A Space Odyssey*, the famous computer HAL is in some ways an Intermediary that mediates between a complex space ship and the astronauts. Although HAL has perfect natural language understanding and advanced commonsense reasoning capabilities, it is not embodied in the physical world, except for its trademark red 'eye' (Figure 90).

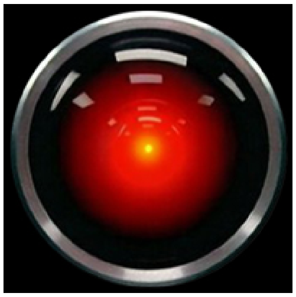


Figure 90: HAL

More recent science fiction is going even further. For example, Gene Roddenberry's TV science fiction show *Andromeda* is about the adventures of a crew on a starship. A special feature of this starship is that it represents itself as an android that inhabits the ship, sharing the same environment as the humanoid crew. This avatar, called "Rommie," embodies an Intermediary in its ultimate perfection.

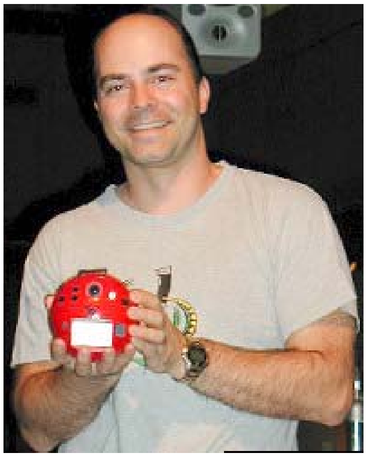


Figure 91: Personal Satellite Assistant

But we do not have to limit ourselves to dreaming about Intermediaries in the domain of space exploration. For some years already, NASA is developing a *Personal Satellite Assistant* (PSA) (Figure 91). The PSA is called an "astronaut support device," designed to move and operate independently in the microgravity environment of space-based vehicles. The PSA will assist astronauts who are living and working aboard the Space Shuttle, Space Station, and during future space exploration missions to the Moon and Mars. The PSA is roughly softball sized and incorporates environmental sensors for gas, temperature, and fire detection, providing the ability for the PSA to monitor spacecraft, payload and crew conditions. Video and audio interfaces support navigation, remote monitoring, and video-conferencing.

In some ways, the PSA is a kind of embodied Intermediary since it inhabits the same area as the astronauts, and can serve as a communication portal to the space ship, other astronauts, and controllers on earth using speech understanding. However, it is still missing non-verbal expressive capabilities like suggested in this thesis.

7.3.5. Some final thoughts

In order to get accepted widely, the Intermediary that I have built may require some social adjustment. Reason is that I suggest that our mobile phones move from *passive* tools and transparent communication portals to proactive Autonomous Intermediaries.

Figure 92: Social adaptation

Will users accept such a change? Will they feel patronized?

Maybe they will feel so at first—but as with many new technologies, people will have to learn to *trust* this novel kind of technology. This may take time, but it's something that we have done many times in the past. For example, when we drive down the highway with 65mph, we trust our car that it won't loose a wheel and kill us all in the following accident. We have learned to trust this technology, and that's why we use it and appreciate its convenience.

Will people accept Intermediaries that act on behalf of them? Will they adjust to such a novel communication paradigms?

Let me put this another way (Figure 92): When phones first got installed in private homes, did people like that they could get interrupted at any time of the day? Certainly not, but they quickly determined that having a phone in their homes has more advantages than disadvantages, and today virtually nobody complains about the presence of phones in homes.

When telephone answering machines were invented, did people like them? No, they hated talking to the 'machine!' These days, people complain when one does not have an answering machine, since we expect to be able to leave messages when nobody picks up the phone.

A few years ago, when somebody walked down the street, talking to himself, people were suspicious and thought that this person was behaving in a strange way, talking to himself—until they learned that he was just talking on his cellphone.

These days, if you walk down the street and talk to yourself, possibly gesturing widely in thin air, all people look for is a thin wire coming out of your ear. If there is such a wire, everything is ok; you are indeed just talking on the phone using a small headset.

It appears like although our social expectations change, we adapt to new social norms.

I am convinced that if keep being open so such changes, and interesting and fascinating future may await us.

8. Epilogue

“The world Lyra lives in is enchanting and intriguing—there, every person, adult and child, has a 'daimon', a sort of animal familiar that, when you're still a child, changes to match your mood. If you're angry, it'll be a wildcat, if you're timid, a mouse, frightened, a bird. Like that. When you become an adult, your daimon turns into whatever you 'really' are the most. It seems to make people more honest, more genuine in Lyra's world than in ours...no matter what the person *says*, you can look to their daimon to see what they're really thinking.

Daimons also talk to their humans, and occasionally to other humans, but the daimon/human bond is one that goes all the way back to the Garden of Eden for this world. The daimons are seen as your soul.

Lyra's daimon is named Pantalaimon, and the bond between the two is fun, as sometimes they scold one another and get into little scuffles...no matter what, though, they're literally inseparable. If they get more than about ten feet apart, they both become weak, and start to wilt (…).” (*The Golden Compass* by Philip Pullman) [161]

9. References

[1] Allen, J. F. (1994). Natural Language Understanding. 2nd edition. The Benjamin/Cummings Publishing Company, Menlo Park, California, (Addison-Wesley Publishing Company, Reading, Massachusetts), 1994, 550 pages, ISBN 0-8053-0330-8.

[2] Aoki, P. M., Romaine, M., Szymanski, M. H. Thornton, J. D., Wilson, D., Woodruff, A. (2003). The Mad Hatter's Cocktail Party: A Mobile Social Audio Space Supporting Multiple Simultaneous Conversations. Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '03), Ft. Lauderdale, FL, April 2003, pp 425-432. http://doi.acm.org/10.1145/642611.642686

[3] Avrahami, D., Gergle, D., Hudson, S.E., Kiesler, S. (2003). Improving the Accuracy of Cell Phone Interruptions: A Study on the Effect of Contextual Information on the Behavior of Callers. Unpublished manuscript. http://www.cs.cmu.edu/~nx6/research/317_Avrahami_submitted_CHI2003.pdf

[4] Ball, G., Ling, D., Kurlander, D., Miller, J., Pugh, D., Skelly, T., Stankosky, A., Theil, D., Van Dantzich, M., & Wax, T. (1997). Lifelike computer characters: The persona project at Microsoft research. In J. M. Bradshaw (Ed.), Software Agents (pp 191-222). Menlo Park, CA: AAAI/MIT Press.

[5] Barkhuus, L. (2003). How to Define the Communication Situation: Determining Context Cues in Mobile Telephony. CONTEXT 2003, pp 411-418. http://www.itu.dk/~barkhuus/context03poster.pdf

[6] Barkhuus, L., Dey, A.K. (2003). Is context-aware computing taking control away from the user? Three levels of interactivity examined. UBICOMP 2003, 5th International Symposium on Ubiquitous Computing, to appear. October 12-15, 2003. http://intel-research.net/Publications/Berkeley/072120031254_134.pdf

[7] Bartneck, C., Okada, M. (2001a), Robofesta - Robotic User Interfaces in Japan. Symposium on Multimodal Communication with Embodied Agents, CWI, Amsterdam [video]

[8] Bartneck, C., Okada, M. (2001b). Robotic User Interfaces. Proceedings of the Human and Computer Conference (HC-2001), Aizu, pp 130-140. http://www.bartneck.de/work/bartneck_HC2001.pdf

[9] Basu, S. (2002). Conversational Scene Analysis. Ph.D. Thesis, MIT Department of EECS, September 2002. http://www.media.mit.edu/~sbasu/papers/thesis.pdf

[10] Bedau, M., McCaskill, J., Packard, N. Rasmussen, S. Adami, C., Green, D., Ikegami, T., Kaneko K., Ray, T. (2000). Open Problems in Artificial Life. Artificial Life 6(4), pp 363-376. http://mitpress.mit.edu/journals/ARTL/Bedau.pdf

[11] Berner, U., Rieger, T., Müller, W. (2000). Conversational Interfaces. COMPUTER GRAPHIK Topics 11/2000 (Special Edition) Reports of the INI-GraphicsNet. http://www.inigraphics.net/publications/topics/2000/SA_2000/SA00_a04.pdf
http://www.igd.fhg.de/~rieger/publications/TopicsSpecial.pdf

[12] Beutel, J., Kasten, O., Ringwald, M., Siegemund, F., Thiele, L. (2003). Bluetooth Smart Nodes for Ad-hoc Networks. TIK Report Nr. 167, ETH Zurich, April, 2003. ftp://ftp.tik.ee.ethz.ch/pub/people/beutel/20030410_testbed_tikreport.pdf

[13] Bickmore, T. (1999). Social Intelligence in Conversational Computer Agents: Proseminar Conceptual Analysis of Thesis Area.

[14] Bickmore, T. (2003). Relational Agents: Effecting Change through Human-Computer Relationships. MIT Ph.D. Thesis, February 2003. http://web.media.mit.edu/~bickmore/bickmore-thesis.pdf

[15] Bickmore, T., Cassell, J. (2000). 'How about this weather?' - Social Dialogue with Embodied Conversational Agents. Proceedings of the AAAI Fall Symposium on Socially Intelligent Agents: The Human in the Loop, November 3-5, North Falmouth, MA, Technical Report FS-00-04, pp 4-8. http://gn.www.media.mit.edu/groups/gn/publications/SIA_2000.pdf

[16] Blinkoff, R. (2003). The Mobiles: Social Evolution in a Wireless Society. Proprietary Ethnographic Research Study. Baltimore, MD: Context-Based Research Group. http://www.contextresearch.com/context/study.cfm

[17] Boyce, S. J. (2000). Natural spoken dialogue systems for telephony applications. Communications of the ACM 43(9), pp 29-34. http://doi.acm.org/10.1145/348941.348974

[18] Breazeal, C. (1999). Robot in Society: Friend or Appliance? In Agents99 Workshop on Emotion-Based Agent Architectures, Seattle, WA, pp 18-26. http://www.ai.mit.edu/projects/kismet/Breazeal-Agents99.ps.gz

[19] Breazeal, C. L. (2002). Designing sociable robots. Cambridge, Mass.: MIT Press.

[20] Breazeal, C., Buchsbaum, D., Gray, J., Gatenby, D., Blumberg, B. (2005). Learning From and About Others: Towards Using Imitation to Bootstrap the Social Understanding of Others by Robots. In L. Rocha and F. Almedia e Costa (eds.), Artificial Life, January 2005, vol. 11, no. 1-2, pp. 31-62. http://robotic.media.mit.edu/Papers/Breazeal-etal-AL04.pdf

[21] Breazeal, C., Fitzpatrick, P. (2000). That Certain Look: Social Amplification of Animate Vision. Proceedings of the AAAI Fall Symposium on Socially Intelligent Agents: The Human in the Loop, November 3-5, North Falmouth, MA, Technical Report FS-00-04, pp 18-23. http://www.ai.mit.edu/people/paulfitz/pub/AAAIFS00.pdf

[22] Breazeal, C., Foerst, A. (1999). Schmoozing with Robots: Exploring the Boundary of the Original Wireless Network. Proceedings of the 1999 Conference on Cognitive Technology (CT99), San Francisco, CA, pp 375-390. http://www.cogtech.org/CT99/breazeal.htm

[23] Brooks, K. (2003). Personal communication, November 23, 2003.

[24] Brooks, R. A. (1999). Cambrian intelligence: the early history of the new AI. Cambridge, Mass.: MIT Press.

[25] Brooks, R. A. (2002). Flesh and machines: how robots will change us (1st ed.). New York: Pantheon Books.

[26] Campbell, C.S., Tarasewich, P. (2004). Designing visual notification cues for mobile devices. Proceedings of CHI'04 Extended Abstracts, pp 1199-1202. http://doi.acm.org/10.1145/985921.986023

[27] Choudhury, T. (2003). Sensing and Modeling Human Networks. Ph.D. Thesis, MIT Media Lab, August 2003.

[28] Choudhury, T., Pentland, A. (2003a). Sensing and Modeling Human Networks using the Sociometer. Submitted to the International Conference on Wearable Computing. October 2003. White Plains, New York. ftp://whitechapel.media.mit.edu/pub/tech-reports/TR-572.pdf

[29] Choudhury, T., Pentland, A. (2003b). Modeling Face-to-Face Communication using the Sociometer. To Appear in: Proceedings of the International Conference on Ubiquitous Computing, Seattle, WA. October 2003. ftp://whitechapel.media.mit.edu/pub/tech-reports/TR-564.pdf

[30] Clarkson, B. (2002). Life Patterns: structure from wearable sensors. Ph.D. Thesis, MIT Media Lab, September 2002. http://web.media.mit.edu/~clarkson/thesis.pdf

[31] Clocksin, W. A. (2000). A Narrative Architecture for Functioning Minds: A Social Constructionist Approach. In Symposium on How to Design a Functioning Mind, AISB2000, University of Birmingham. http://www.clocksin.com/publications/aisb2000.pdf

[32] Cooper, G. (2001). The Mutable Mobile: Social Theory in the Wireless World, in B. Brown, N. Green & R. Harper (eds.), 'Wireless World', Springer, London

[33] Dabbish, L. A., and Baker, R. S. (2003). Administrative assistants as interruption mediators. In Proceedings of ACM Conference on Human Factors in Computing Systems (CHI'03): Extended abstracts. New York: ACM Press, pp 1020-1021. http://doi.acm.org/10.1145/765891.766127

[34] Dautenhahn, K. (1998). The Art of Designing Socially Intelligent Agents - Science, Fiction, and the Human in the Loop. Special Issue Socially Intelligent Agents, Applied Artificial Intelligence Journal, Vol. 12, 7-8, pp 573-617. http://homepages.feis.herts.ac.uk/~comqkd/aain3.ps

[35] Dautenhahn, K. (1999). Embodiment and Interaction in Socially Intelligent Life-Like Agents. In C. L. Nehaniv (ed.) Computation for Metaphors, Analogy and Agent, Springer Lecture Notes in Artificial Intelligence, Volume 1562, New York, NY: Springer, pp 102-142. http://www.springerlink.com/link.asp?id=9m9h2e7ejahq42ur

[36] Dautenhahn, K. (1999b). Robots as Social Actors: AURORA and the Case of Autism. Proceedings of CT99, The Third International Cognitive Technology Conference, August 1999, San Francisco, CA, pp 359-374 http://www.cogtech.org/CT99/Dautenhahn.htm

[37] Dautenhahn, K. (2000). Human cognition and social agent technology. Amsterdam; Philadelphia: John Benjamins.

[38] Dautenhahn, K. (2000b). Socially Intelligent Agents and The Primate Social Brain - Towards a Science of Social Minds. Proceedings of the AAAI Fall Symposium on Socially Intelligent

Agents: The Human in the Loop, November 3-5, North Falmouth, MA, Technical Report FS-00-04, pp 35-51 http://homepages.feis.herts.ac.uk/~comqkd/siaweb00.ps

[39] Dautenhahn, K., Ogden, B., Quick, T. (2002). From embodied to socially embedded agents—implications for interaction-aware robots. Cognitive Systems Research 3(3), pp 397-428. http://homepages.feis.herts.ac.uk/~comqkd/CSR2002.pdf

[40] Davis, D.N. (2000). Minds have personalities - Emotion is the core. In Symposium on How to Design a Functioning Mind, AISB2000, University of Birmingham. http://www.cs.bham.ac.uk/~axs/aisb/papers/davis.pdf

[41] De Haan, G. (1999). The Usability of Interacting with the Virtual and the Real in COMRIS. In: Nijholt, A., Donk, O., and Van Dijk, B. (eds.), Proceedings of TWLT 15 - Interactions in Virtual Worlds, Enschede, The Netherlands, May 19-21, 1999, pp 69-79. http://arti.vub.ac.be/~comris/refmat/Int_Virt_world.html

[42] Dennett, D. C. (1998). Brainchildren: essays on designing minds. Cambridge, Mass.: MIT Press.

[43] Devaul, R., Gips, J., Sung, M. (2003). MIThril 2003: Applications and Architecture. MIT Media Lab Tech Report.

[44] Dryer, D.C., Eisbach, C., Ark, W.S. (1999). At what cost pervasive? A social computing view of mobile computing systems. IBM Systems Journal, Volume 38, No 4. http://researchweb.watson.ibm.com/journal/sj/384/dryer.pdf

[45] Dubberly, H., Mitsch, D. (1987). Knowledge navigator (videotape). ACM SIGGRAPH Video Review 79 (published 1992, tape made in 1987).

[46] Duffy, B.R. (2000). The Social Robot. Ph.D. Thesis, November 2000, Department of Computer Science, University College Dublin, 2000. http://www.medialabeurope.org/people/b-duffy/pubs/BrianDuffy-PhD-Social-Robot.zip

[47] Duffy, B.R. (2003). Anthropomorphism and The Social Robot. Special Issue on Socially Interactive Robots, Robotics and Autonomous Systems 42 (3-4), 31 March 2003. http://vrai-group.epfl.ch/projects/iros2002/papers/4-duffy.pdf

[48] Duncan, S. (1974). On the structure of speaker-auditor interaction during speaking turns. Language in Society 3: pp 161-180.

[49] Eagle, N., Pentland, A. (2003a). Handhelds that Listen and Learn. IEEE Computer Magazine, Special Issue on Handheld Computing, September 2003. http://cba.media.mit.edu/publications/articles/03.09.eagle.ieee2.pdf

[50] Eagle, N., Pentland, A. (2003b). Social Network Computing. Fifth International Conference on Ubiquitous Computing, October 2003. http://cba.media.mit.edu/publications/articles/03.10.eagle.ubicomp.pdf

[51] Eagle, N., Pentland, A. (2003c). The Relationship Barometer: Mobile Phone Therapy for Couples. IEEE Computer Magazine, Special Issue on Handheld Computing, September 2003. http://cba.media.mit.edu/publications/articles/03.09.eagle.ieee.pdf

[52] Eagle, N., Singh, P., Pentland, A. (2003d). Common Sense Conversations: Understanding Casual Conversation using a

Common Sense Database. Artificial Intelligence, Information Access, and Mobile Computing Workshop at the 18th International Joint Conference on Artificial Intelligence (IJAI), August 2003. http://cba.media.mit.edu/publications/articles/03.08.eagle.ijcai.pdf

[53] Edmonds, B. (1998). Modeling Socially Intelligent Agents. Applied Artificial Intelligence, 12:677-699. http://cfpm.org/cpmrep26.html

[54] Edmonds, B., Dautenhahn, K. (1998). The Contribution of Society to the Construction of Individual Intelligence. In Edmonds, B. and Dautenhahn, K. (eds.), Socially Situated Intelligence: a workshop held at SAB'98, August 1998, Zürich. University of Zürich Technical Report, pp 42-60. http://cogprints.ecs.soton.ac.uk/archive/00000802/00/edmonds.pdf

[55] Egbert, M.M. (1997). Schisming: The Collaborative Transformation from a Single Conversation to Multiple Conversations. Research on Language & Social Interaction 30, pp 1-51.

[56] Erickson, T. (2001). Ask not for whom the cell phone tolls: Some problems with the notion of context-aware computing. Communications of the ACM 45(2), February 2002, pp 102-104. http://doi.acm.org/10.1145/503124.503154

[57] Fong, T., Nourbakhsh, I., Dautenhahn, K. (2002). A Survey of Socially Interactive Robots: Concepts, Design, and Applications. Technical report CMU-RI-TR-02-29, Robotics Institute, Carnegie Mellon University, December 2002. http://www.ri.cmu.edu/pub_files/pub3/fong_terrence_w_2002_3/fong_terrence_w_2002_3.pdf

[58] Fong, T., Nourbakhsh, I., Dautenhahn, K. (2003). A Survey of Socially Interactive Robots. Robotics and Autonomous Systems, vol. 42(3-4), March 2003. http://www.ri.cmu.edu/pub_files/pub3/fong_terrence_w_2003_4/fong_terrence_w_2003_4.pdf

[59] Ford, K. M., Glymour, C. N., & Hayes, P. J. (1995). Android epistemology. Menlo Park, Cambridge, Mass.: AAAI Press; MIT Press.

[60] Franklin, S. (1995). Artificial minds. Cambridge, Mass.: MIT Press.

[61] Fraunhofer Institut, Software- und Systemtechnik: Der Digitale Begleiter. Informationslogistik News Nr. 17 (July 2001). http://www.informationslogistik.org/newsletter/pdf/newsletter17.pdf http://www.informationslogistik.org/publikationen/pdf/digitaler-kumpel-prod-bl_neu.pdf

[62] Geldof, S. (1999). Parrot-talk Requires Multiple Context Dimensions. In: P. Brézillon and P. Bouquet (Eds.) Proceedings of the 2nd Intl. and Interdisciplinary Conference on Modeling and Using Context, Trento (I) Sept. 1999, LNAI Volume 1688, Heidelberg: Springer Verlag, pp 467-470. http://www.springerlink.com/openurl.asp?genre=article&issn=0302-9743&volume=1688&spage=467

[63] Geldof, S., Terken, J. (2001). Talking Wearables Exploit Context. Personal and Ubiquitous Computing 5(1), Jan. 2001, pp 62-65. http://dx.doi.org/10.1007/s007790170033

[64] Gerasimov, V. (2001). Hoarder board. http://vadim.www.media.mit.edu/Hoarder/Hoarder.htm

[65] Gerasimov, V., Selker, T., Bender, W. (2002). Sensing and effecting environment with extremity-computing devices. Motorola OFFSPRING, VOL 1, NO 1, 2002, pp 1-9. http://vadim.www.media.mit.edu/Temp/Extremity.pdf

[66] Geser, H. (2002). Towards a Sociological Theory of the Mobile Phone. University of Zürich, Switzerland, August 2002. http://socio.ch/mobile/t_geser1.htm

[67] Goffman, E. (1966). Alienation from Interaction. Communication and Culture, Alfred G Smith (ed.) Holt, Rinehart and Winston, New York, 1966.

[68] Gong, L. (2002). Towards a Theory of Social Intelligence for Interface Agents. Paper presented at workshop Virtual Conversational Characters: Applications, Methods, and Research Challenges, 29th November 2002, Melbourne, Australia. http://www.vhml.org/workshops/HF2002/papers/gong/gong.pdf

[69] Gottlieb, H. (1997). An Outline of The Jack Principles of the Interactive Conversation Interface (ICI), The Short Version. Unpublished white paper, first edition 1997-2002, Jellyvision Inc.

[70] Gottlieb, H. (2002). The interactive conversation interface (ICI): a proposed successor to GUI for an interactive broadband world. Proceedings of the 2002 International Conference on Intelligent User Interfaces (IUI 2002), January 13-16, San Francisco, California, USA, p 2. http://doi.acm.org/10.1145/502716.502718

[71] Graham, J. (2003). Hello, tech designers? This stuff is too small. USA TODAY, March 3, 2003. http://www.usatoday.com/tech/news/techinnovations/2003-03-03-tiny-tech_x.htm

[72] Green, A. (2001). C-Roids: Life-like Characters for Situated Natural Language User Interfaces. Proceedings of Ro-Man'01 (10th IEEE International Workshop on Robot and Human Communication), pp140-145, Bordeaux-Paris, September 2001. ftp://ftp.nada.kth.se/IPLab/TechReports/IPLab-193.pdf

[73] Greenberg, S., Fitchett, C. (2001). Phidgets: Easy Development of Physical Interfaces through Physical Widgets. In Proceedings of ACM UIST'01, pp 209-218. http://doi.acm.org/10.1145/502348.502388

[74] Greenberg, S., Kuzuoka, H. (2000). Using Digital but Physical Surrogates to Mediate Awareness, Communication and Privacy in Media Spaces. Personal Technologies, 4(1), January. http://www.cpsc.ucalgary.ca/grouplab/papers/2000/00-Surrogate.PersTechnology/dig-phys-perstech.pdf

[75] Haddon, L. (2000) The Social Consequences of Mobile Telephony: Framing Questions. Paper presented at the seminar 'Sosiale Konsekvenser av Mobiltelefoni', organised by Telenor, 16th June, Oslo. http://members.aol.com/leshaddon/Framing.html

[76] Hansson, R., Ljungstrand, P. (2000). The Reminder Bracelet: Subtle Notification Cues for Mobile Devices. Extended Abstracts of CHI 2000 (Student Poster), ACM Press, pp 323-325. http://doi.acm.org/10.1145/633292.633488

[77] Hansson, R., Ljungstrand, P., Redström, J. (2001). Subtle and Public Notification Cues for Mobile Devices. Proceedings of UbiComp 2001, Atlanta, Georgia, USA. http://www.viktoria.se/~rebecca/artiklar/Notification_final.pdf

[78] Hill, J. L. (2003). System Architecture for Wireless Sensor Networks. Ph.D. Thesis, University of California, Berkeley. http://www.cs.berkeley.edu/~jhill/jhill_thesis.pdf

[79] Hinckley, K., Horvitz, E. (2001). Toward More Sensitive Mobile Phones. ACM UIST 2001 Symposium on User Interface Software & Technology, pp. 191-192. http://research.microsoft.com/users/kenh/papers/MobilePhoneSensing_UIST01.pdf

[80] Hofstadter, D. R., Dennett, D. C. (1981). The mind's I: fantasies and reflections on self and soul. New York: Basic Books.

[81] Hogg, L. M. J., Jennings, N. R. (1997). Socially Rational Agents. AAAI Fall Symposium on Social Intelligent Agents, pp 61-63. http://eprints.ecs.soton.ac.uk/archive/00002154/

[82] Hogg, L. M. J., Jennings, N. R. (2001). Socially intelligent reasoning for autonomous agents. IEEE Trans on Systems, Man and Cybernetics - Part A 31(5):381-399. http://www.ecs.soton.ac.uk/~nrj/download-files/smc-01.pdf

[83] Holmquist, L.E., Mattern, F., Schiele, B., Alahuhta, P., Beigl, M., Gellersen, H. (2001). Smart-Its Friends: A Technique for Users to Easily Establish Connections between Smart Artefacts. Proceedings of Ubicomp 2001, Atlanta, USA, October 2001. http://www.inf.ethz.ch/vs/publ/papers/smf.pdf

[84] Horvitz, E., Jacobs, A., Hoxel, D. (1999). Attention-sensitive alerting. Proceedings of UAI '99, Conference on Uncertainty and Artificial Intelligence (UAI), Stockholm, Sweden, July 1999. Morgan Kaufmann: San Francisco. pp 305-313. ftp://ftp.research.microsoft.com/pub/ejh/priorities.pdf http://doi.acm.org/10.1145/792704.792733

[85] Hudson, S.E, Fogarty, J., Atkeson, C.G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J.C., Yang, J. (2003). Predicting Human Interruptibility with Sensors: A Wizard of Oz Feasibility Study. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2003). http://doi.acm.org/10.1145/642611.642657

[86] IDEO: Social Mobiles. http://www.ideo.com/case_studies/Social_Mobiles/index.html

[87] Intille, S. S. (2002). Change Blind Information Display for Ubiquitous Computing Environment. Proceedings of UbiComp 2002, pp 91-106. http://www.media.mit.edu/~intille/papers-files/ubicomp02.pdf

[88] Isaacs, E., Walendowski, A., Ranganathan, D. (2002). Hubbub: A sound-enhanced mobile instant messenger that supports awareness and opportunistic interactions. In Proceedings of ACM CHI '02), Minneapolis, MN, pp 179-186. http://doi.acm.org/10.1145/503376.503409

[89] Isbister, K. (2003). Social Signals: Using Principles and Methods from Social Psychology to Guide Subtle Expression Design. Proceedings of the CHI 2003 Workshop on Subtle Expressivity for Characters and Robots. April 7th, Fort Lauderdale, FL.

http://www.mis.atr.co.jp/~noriko/CHI2003/proceedingsCHI03Wrkshp.pdf

[90] Jabarin, B., Wu, J., Vertegaal, R., Grigorov, L. (2003). Establishing Remote Conversations Through Eye Contact With Physical Awareness Proxies. In Extended Abstracts of ACM CHI 2003 Conference on Human Factors in Computing Systems, pp 948-949. http://doi.acm.org/10.1145/765891.766087

[91] Kahn, J. M., Katz, R. H., Pister, K. S. J. (1999). Next century challenges: mobile networking for "Smart Dust". Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking (MobiCom 99), Seattle, WA, August 17-19, pp 271-278. http://doi.acm.org/10.1145/313451.313558

[92] Kaminsky, M., Dourish, P., Edwards, K. LaMarca, A., Salisbury, M., Smith, I. (1999). SWEETPEA: Software tools for programmable embodied agents. Proceedings of ACM CHI 99 Conference on Human Factors in Computing Systems, pp 144-151. http://doi.acm.org/10.1145/302979.303021

[93] Kasten, O., Langheinrich, M. (2001). First Experiences With Bluetooth in the Smart-Its Distributed Sensor Network. Proceedings of the Workshop on Ubiquitous Computing and Communications (PACT), Oct. 2001. http://www.inf.ethz.ch/vs/publ/papers/bt-experiences.pdf

[94] Katz, J., Aakhus, M. (Eds.) (2000). Perpetual contact: Mobile communication, private talk, public performance. Cambridge: Cambridge University Press.

[95] Kidd, C., & Breazeal, C. (2004). Effect of a Robot on Engagement and User Perceptions. Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS04), Sendai, Japan. http://robotic.media.mit.edu/Papers/Kidd_Iros04.pdf

[96] Kidd, C.D., O'Connell, T., Nagel, K., Patil, S., Abowd, G.D. (2000). Building a Better Intercom: Context-Mediated Communication within the Home. Georgia Tech GVU Tech Report #00-27. http://web.media.mit.edu/~coryk/papers/00-27.pdf

[97] Kidder, L. M. (1981). Research Methods in Social Relations. New York: Holt, Rinehart & Winston.

[98] Kihlstrom, J.F., Cantor, N. (2000). Social intelligence. In R.J. Sternberg (Ed.), Handbook of intelligence, 2nd ed. (pp 359-379). Cambridge, U.K.: Cambridge University Press. http://socrates.berkeley.edu/~kihlstrm/social_intelligence.htm

[99] Kim, H., Chan, P. (2003). Learning Implicit User Interest Hierarchy for Context in Personalization. Proceedings of International Conference on Intelligent User Interfaces (IUI), pp 101-108. http://www.iuiconf.org/03pdf/2003-001-0033.pdf

[100] Kortuem, G., Schneider, J., Preuitt, D., Cowan Thompson, T.G., Fickas, S., Segall, Z. (2001). When Peer-to-Peer comes Face-to-Face: Collaborative Peer-to-Peer Computing in Mobile Ad-hoc Networks. Proceedings 2001 International Conference on Peer-to-Peer Computing (P2P2001), Aug 27-29, 2001, Linköping, Sweden. http://www.cs.uoregon.edu/research/wearables/Papers/p2p2001.pdf

[101] Kurzweil, R. (1999). The age of spiritual machines: when computers exceed human intelligence. New York: Viking.

[102] Kuzuoka, H., Greenberg, S. (1999). Mediating Awareness and Communication through Digital but Physical Surrogates. Proceedings of CHI'99 Extended Abstracts, pp 11-12. http://doi.acm.org/10.1145/632716.632725

[103] Kuzuoka, H., Ishimoda, G., Nishimura, Y., Suzuki, R., Kondo, K. (1995). Can the GestureCam be a Surrogate? Proceedings of ECSCW'95, 1995, pp 181-196. http://www.kuzuoka-lab.esys.tsukuba.ac.jp/lab/gesturecam/gesturecam.html

[104] Lashkari, Y., Metral, M., Maes, P. (1994). Collaborative Interface Agents. Proceedings of AAAI '94 Conference, Seattle, Washington, August 1994. http://agents.www.media.mit.edu/groups/agents/publications/aaai-ymp/aaai.html

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

[106] Lee, S. -I., Sung, C., Cho, S.-B. (2001). An effective conversational agent with user modeling based on bayesian network. Web Intelligence 2001, pp 28-432. http://candy.yonsei.ac.kr/publications/Papers/AECAUMBOBN.pdf

[107] Lenat, D. B., Guha, R. V. (1989). Building large knowledge-based systems: representation and inference in the Cyc project. Reading, Mass.: Addison-Wesley Pub. Co.

[108] Licklider, J.C.R., Taylor, R. (1968). The computer as a communication device. Science & Technology, 76, pp 21-31. http://gatekeeper.dec.com/pub/DEC/SRC/publications/taylor/licklider-taylor.pdf

[109] Lieberman, H., Liu, H., Singh, P., Barry, B. (2003). Beating Some Common Sense into Interactive Applications. Paper submitted to the International Joint Conference on Artificial Intelligence (IJCAI) 2003 (draft). http://www.media.mit.edu/~push/Beating-Common-Sense.pdf

[110] Lieberman, H., Selker, T. (2000). Out of context: Computer systems that adapt to, and learn from, context. IBM Systems J. 39, 3&4, pp 617–632. http://www.research.ibm.com/journal/sj/393/part1/lieberman.html

[111] Liechti, O., Sifer, N., Ichikawa, T. (1999). A Non-obtrusive User Interface for Increasing Social Awareness on the World Wide Web. Personal Technologies 3 (1&2), 1999: 22-32. http://www.mic.atr.co.jp/~olivier/papers/liechti-awareness-2.pdf

[112] Lifton, D. Seetharam, M. Broxton, J. Paradiso (2002). Pushpin Computing System Overview: a Platform for Distributed, Embedded, Ubiquitous Sensor Networks. In F. Mattern and M. Naghshineh (eds): Pervasive 2002, Proceedings of the Pervasive Computing Conference, Zurich Switzerland, 26-28 August 2002, Springer Verlag, Berlin Heidelberg, pp 139-151. http://www.media.mit.edu/resenv/pubs/papers/2002-09-PushpinPervasiveWF.pdf

[113] Ling, R. (1997). ‘One can talk about common manners!’: the use of mobile telephones in inappropriate situations. In L. Haddon (ed.): Themes in mobile telephony, Final Report of the COST 248 Home and Work group. http://www.telenor.no/fou/program/nomadiske/articles/09.pdf

[114] Ling, R. (2000). Direct and mediated interaction in the maintenance of social relationships. In Sloane, A. and van Rijn, F. (eds.) Home informatics and telematics: Information, technology and society. Kluwer, Boston, pp 61-86. http://www.telenor.no/fou/program/nomadiske/articles/02.pdf

[115] Liu, H., Lieberman, H., Selker, T. (2002). GOOSE: A Goal-Oriented Search Engine With Commonsense. In De Bra, Brusilovsky, Conejo (Eds.): Adaptive Hypermedia and Adaptive Web-Based Systems, Second International Conference, AH 2002, Malaga, Spain, May 29-31, 2002, Proceedings. Lecture Notes in Computer Science 2347 Springer 2002, ISBN 3-540-43737-1, pp 253-263. http://web.media.mit.edu/~hugo/publications/papers/AH2002-goose.pdf

[116] Liu, H., Lieberman, H., Selker, T. (2003). A Model of Textual Affect Sensing using Real-World Knowledge. Proceedings of the Seventh International Conference on Intelligent User Interfaces (IUI 2003), Miami, Florida, pp 125-132. http://web.media.mit.edu/~hugo/publications/papers/IUI2003-affectsensing.pdf

[117] Liu, H., Singh. P. (2004). ConceptNet: A Practical Commonsense Reasoning Toolkit. BT Technology Journal 22(4). Kluver, Boston, pp 211-226. http://web.media.mit.edu/~hugo/publications/papers/BTTJ-ConceptNet.pdf

[118] Liu, K.K., Picard, R.W. (2003). Subtle Expressivity in a Robotic Computer. Proceedings of the CHI 2003 Workshop on Subtle Expressivity for Characters and Robots. April 7th, Fort Lauderdale, FL. http://web.media.mit.edu/~kkliu/publications/roco_final.pdf

[119] Lockerd, A.L. (2002a). DriftCatcher: Understanding Implicit Social Context in Electronic Communication. MIT Master's Thesis, September 2002. http://www.media.mit.edu/~alockerd/andrea-final.pdf

[120] Lockerd, A.L. (2003). DriftCatcher: The Implicit Social Context of Email. In proceedings of the Ninth IFIP TC13 International Conference on Human-Computer Interaction (INTERACT 2003), Sep. 1-5, 2003, Zuerich, Switzerland.

[121] Lockerd, A.L., Selker, T. (2002b). DriftCatcher: Enhancing Social Networks Through Email. Presented at SUNBELT XXII International Sunbelt Social Network Conference, 13-17 February 2002, Le Meridien, New Orleans, USA. http://cac.media.mit.edu:8080/contextweb/sunbelt_paper.pdf

[122] Love, S., Perry, M. (2004). Dealing with mobile conversations in public places: some implications for the design of socially intrusive technologies. Extended Abstracts of CHI 2004, pp 1195-1198. http://doi.acm.org/10.1145/985921.986022

[123] Madden, S., Franklin, M., Hellerstein, J., Hong, W. (2002). TAG: A Tiny Aggregation Service for Ad-hoc Sensor Networks. Symposium on Operating Systems Design and Implementation (OSDI 2002), Boston, December 2002. http://www.cs.berkeley.edu/~madden/madden_tag.pdf

[124] Maes, P. (1994). Agents that reduce work and information overload. Communications of the ACM 37(7), July 1994, pp 30-40. http://doi.acm.org/10.1145/176789.176792

[125] Maglio, P., Matlock, T,. Campbell, C., Zhai, S., Smith, B.A. (2000). Gaze and speech in attentive user interfaces. Proceedings of The third International Conference on Multimodal Interfaces, Oct 14-16, 2000, Beijing, China. http://www.almaden.ibm.com/cs/people/pmaglio/pubs/mmui5.pdf

[126] Marcus, A. (2002) The cult of cute: the challenge of user experience design. ACM Interactions Nov/Dec, pp 29-34. http://doi.acm.org/10.1145/581951.581966

[127] Marti, S., Schmandt, C. (2005). Giving the Caller the Finger: Collaborative Responsibility for Cellphone Interruptions. Extended Abstracts of CHI2005, pp 1633-1636. http://doi.acm.org/10.1145/1056808.1056984

[128] Marx, M., Schmandt, C. (1996). CLUES: Dynamic Personalized Message Filtering. Proceedings of CSCW '96, November 1996, pp 113-121. http://www.media.mit.edu/speech/papers/1996/marx_CSCW96_clues.pdf

[129] Maybury, M.T., Wahlster, W. (eds.) (1997). Readings in Intelligent User Interfaces. Menlo Park, CA: Morgan Kaufmann.

[130] McBean, J., & Breazeal, C. (2004). Voice Coil Actuators for Human-Robot Interaction. Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS04), Sendai, Japan. http://robotic.media.mit.edu/Papers/McBean-Iros04.pdf

[131] McFarlane, D. C. (1997). Interruption of People in Human-Computer Interaction: A General Unifying Definition of Human Interruption and Taxonomy. NRL Formal Report NRL/FR/5510-97-9870, Washington: US Naval Research Laboratory. http://interruptions.net/literature/McFarlane-NRL-97.pdf

[132] McLuhan, M. (1964). Understanding Media. New York: McGraw-Hill.

[133] McTear , M. (2002). Spoken dialogue technology: enabling the conversational user interface. ACM Computing Surveys (CSUR), 34(1), March 2002, pp 90-169. http://doi.acm.org/10.1145/505282.505285

[134] Menzel, P., D'Aluisio, F., Mann, C. C. (2000). Robo sapiens: evolution of a new species. Cambridge, Mass.: MIT Press.

[135] Michaud, F. (2000). Social intelligence and robotics. Proceedings of the AAAI Fall Symposium on Socially Intelligent Agents: The Human in the Loop, November 3-5, North Falmouth, MA, Technical Report FS-00-04, pp 127-130. http://www.gel.usherb.ca/laborius/papers/AAAIf00.pdf

[136] Milewski, A., Smith, T. M. (2000). Providing Presence Cues to Telephone Users. Proceedings of CSCW 2000, Philadelphia, PA.

http://www.research.att.com/~tsmith/papers/milewski-smith.pdf

[137] Miller, C. (2000). Rules of Etiquette, or How a Mannerly AUI should Comport Itself to Gain Social Acceptance and be Perceived as Gracious and Well-Behaved in Polite Society. In the Working Notes of the AAAI-Spring Symposium on Adaptive User Interfaces, March 20-22, 2000, Stanford, CA. http://www.sift.info/English/publications/AAAI-SS-00.pdf

[138] Miller, G. A. (1995). WordNet: a lexical database for English. Communications of the ACM 38(11), November 1995, pp 39-41. http://doi.acm.org/10.1145/219717.219748

[139] Miner, C.S. (2001). Pushing functionality into even smaller devices. Communications of the ACM, 44(3), March 2001, pp 72-73. http://doi.acm.org/10.1145/365181.365205

[140] Miner, C.S., Chan, D.M., & Campbell, C.S. (2001). Digital Jewelry: Wearable Technology for Everyday Life. In Proceedings of the Conference on Human Factors in Computing Systems, Extended Abstracts (CHI 2001), pp 45-46. http://doi.acm.org/10.1145/634067.634098

[141] Minksy, M., Singh, P. (2002). The St. Thomas Commonsense Symposium. Revised version to appear in AI Magazine. http://www.media.mit.edu/~push/StThomasSymposium.pdf

[142] Minsky, M. (2000). Deep issues: commonsense-based interfaces. Communications of the ACM 43(8), August 2000, pp 66-73. http://doi.acm.org/10.1145/345124.345145

[143] Minsky, M. L. (1986). The society of mind. New York, N.Y.: Simon and Schuster.

[144] Moravec, H. (1999). Robot: Mere machine to transcendent mind. New York: Oxford University Press.

[145] Mynatt, E., Back, M., Want, R., Baer, M., Ellis, J. (1998). Designing audio aura. Proceedings of CHI '98, pp 566-573. http://doi.acm.org/10.1145/274644.274720

[146] Nagel, K., Kidd, C.D., O'Connell, T., Dey, A.K., Abowd, G.D. (2001). The Family Intercom: Developing a Context-Aware Audio Communication System. Proceedings of UBICOMP 2001, September 30-October 2, 2001, pp 176-183. http://www.cc.gatech.edu/~kris/publications/pdf/ubi01.pdf

[147] Nelson, L., Bly, S., Sokoler T. (2001). Interactive Quiet Calls: Talking Silently on Mobile Phones. Proceedings of CHI 2001, pp 171-181. http://doi.acm.org/10.1145/365024.365094

[148] Nicolescu, M.N., Mataric, M.J. (2001). Learning and Interacting in Human-Robot Domains. IEEE Transactions on Systems, Man and Cybernetics, Part A, vol. 31, n.5, pp 419-430. http://icat.or.kr/robotics/PDF/1_15n/Learning.pdf

[149] Norman, D. A. (1998). The invisible computer: why good products can fail, the personal computer is so complex, and information appliances are the solution. Cambridge, Mass.: MIT Press.

[150] Norman, D. A. (2004). Emotional design: why we love (or hate) everyday things. New York: Basic Books.

[151] Okada, M., Suzuki, N., Date, M. (1999). Social Bonding in Talking with Social Autonomous Creatures. Proceedings of EuroSpeech-99, S9.OR2.4, pp 1731-1734.

http://www.telecom.tuc.gr/paperdb/eurospeech99/PAPERS/S9O2/O025.PDF

[152] Osgood, E.C., Suci, G.J., Tannenbaum, P.H. (1957). The measurement of meaning. Urbana: University of Illinois Press.

[153] Pedersen, E., Sokoler, T. (1997). AROMA: Abstract Representation of Presence Supporting Mutual Awareness. Proceedings of CHI 1997, pp 51-58. http://doi.acm.org/10.1145/258549.258584

[154] Pedersen, E.R. (2001). Calls.calm: Enabling Caller and Callee to Collaborate. Extended Proceedings of CHI '01, Seattle, pp 235-236. http://doi.acm.org/10.1145/634067.634207

[155] Pering, C. (2002). Interaction Design Prototyping of Communicator Devices: Towards Meeting the Hardware-Software Challenge. ACM Interactions, 9 (6), pp 36-46. http://doi.acm.org/10.1145/581951.581952

[156] Pering, C. (2002). Taming of the ring: context specific social mediation for communication devices. Extended Proceedings of CHI 2002, pp 712-713. http://doi.acm.org/10.1145/506443.506560

[157] Plant, S. (2000). On the Mobile: The effect of mobile telephones on social and individual life. http://www.motorola.com/mot/documents/0,,296,00.pdf http://www.receiver.vodafone.com/06/articles/pdf/01.pdf

[158] Poor, R. (2001). Embedded Networks: Pervasive, Low-Power, Wireless Connectivity. Ph.D. thesis, Program in Media Arts and Sciences, Massachusetts Institute of Technology, May 2001.

[159] Poor, R., Hodges, B. (2002). Reliable Wireless Networks for Industrial Systems. Technical White Paper, Ember Corporation.

[160] Poupyrev I., Maruyama S., Rekimoto J. (2002). Ambient Touch: Designing Tactile Interfaces for Handheld Devices. Proceedings of UIST 2002, Paris, France, pp 51-60. http://doi.acm.org/10.1145/571985.571993

[161] Pullman, P. (1995). Northern Lights (The Golden Compass). Scholastic Press, ISBN 0590541781. http://www.randomhouse.com/features/pullman/goldencompass/

[162] Pynadath, D., Tambe, M., Arens, Y., Chalupsky, H., Gil, Y., Knoblock, C., Lee, H., Lerman, K., Oh, J., Ramachandran, S., Rosenbloom, P.S., Russ, T. (2000). Electric Elves: Immersing an agent organization in a human organization. Proceedings of the AAAI Fall Symposium on Socially Intelligent Agents: The Human in the Loop, November 3-5, North Falmouth, MA, Technical Report FS-00-04, pp 150-154. http://www.isi.edu/teamcore/tambe/papers/2000/elves.ps

[163] Rao, S. Garg, S., Lieberman, H., Liu, H. (2002). Commonsense via Instant Messaging. MIT Media Lab, Technical report. http://ocw2.mit.edu/NR/rdonlyres/5C7CF160-84C0-4E56-BD32-5F36B6CCE2AE/0/proj_file9_wor.pdf

[164] Reeves, B., Nass, C. I. (1996). The media equation: how people treat computers, televisions, and new media like real people and places. Stanford, Calif. New York: CSLI Publications; Cambridge University Press.

[165] Rhee, S., Liu, S. (2002). An Ultra-low Power, Self-Organizing Wireless Network and Its Applications to Non-invasive

Biomedical Instrumentation. IEEE/Sarnoff Symposium on Advances in Wired and Wireless Communications, West Trenton, New Jersey, March 13, 2002. http://www.mit.edu/people/sokwoo/20020222_SarnoffSymposium2002Paper.pdf

[166] Rhee, S., Seetharam, D., Liu, S., Wang, N., Xiao, J. (2003). i-Beans: An Ultra-low Power Wireless Sensor Network. UBICOMP 2003, 5th International Symposium on Ubiquitous Computing, October 12-15, 2003.

[167] Rheingold, H. (2002). Smart Mobs: The Next Social Revolution. Cambridge: Perseus.

[168] Rich, C., Lesh, N.B., Rickel, J., Garland, A. (2002). A Plug-in Architecture for Generating Collaborative Agent Responses. International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), ISBN: 1-58113-480-0, pp 782-789. http://doi.acm.org/10.1145/544862.544924

[169] Riley, P. (1976). Discursive and Communicative Functions of Non-Verbal Communication. ED143217/FL008785, ERIC (Educational Resources Information Center), Sewell, NJ.

[170] Sawhney, N., Schmandt, C. (1999). Nomadic Radio: Scalable and contextual notification for wearable audio messaging. Proceedings of CHI'99, pp 96-103. http://doi.acm.org/10.1145/302979.303005

[171] Sawhney, N., Schmandt, C. (2000). Nomadic Radio: Speech & Audio Interaction for Contextual Messaging in Nomadic Environments. ACM Transactions on CHI, vol. 7, pp 353-383. http://doi.acm.org/10.1145/355324.355327

[172] Scassellati, B. (2000). Theory of Mind... for a Robot. Proceedings of the AAAI Fall Symposium on Socially Intelligent Agents: The Human in the Loop, November 3-5, North Falmouth, MA, Technical Report FS-00-04, pp 164-168. http://www.ai.mit.edu/projects/lbr/hrg/2000/Scassellati-SIA.pdf

[173] Scerri, P., Tambe, M., Lee, H., Pynadath, D. (2000). Don't cancel my Barcelona trip: Adjusting the autonomy of agent proxies in human organizations. Proceedings of the AAAI Fall Symposium on Socially Intelligent Agents: The Human in the Loop, November 3-5, North Falmouth, MA, Technical Report FS-00-04, pp 169-173. http://www.isi.edu/teamcore/tambe/papers/2000/aa-barcelona.ps

[174] Schmandt, C. (1994). Voice communication with computers: conversational systems. New York: Van Nostrand Reinhold.

[175] Schmandt, C., Arons, B. (1984a). A Conversational Telephone Messaging System. IEEE Transactions on Consumer Electronics CE-30, 3 (August 1984), pp 21-24. http://www.media.mit.edu/speech/papers/1984/schmandt_CE84_conversational_telephone_messaging_system.pdf

[176] Schmandt, C., Arons, B. (1984b). Phone Slave: A Graphical Telecommunications Interface. Proceedings of the Society for Information Display 26, 1 (1984), pp 79-82. http://www.media.mit.edu/speech/papers/1984/schmandt_SID84_phone_slave.pdf

[177] Schmandt, C., Arons, B. (1986). A Robust Parser and Dialog Generator for a Conversational Office System. Proceedings of

the 1986 Conference, pp 355-365. San Jose, CA: American Voice I/O Society (AVIOS), September 1986. http://www.media.mit.edu/speech/papers/1986/schmandt_AVIOS86_robust_parser_and_dialog_generator.pdf

[178] Schmandt, C., Arons, B., and Simmons, C. (1985). Voice Interaction in an Integrated Office and Telecommunications Environment. In Proceedings of 1985 Conference of American Voice I/O Society (AVIOS). http://www.media.mit.edu/speech/papers/1985/schmandt_AVIOS85_voice_interaction_in_office.pdf

[179] Schmandt, C., Marmasse, N., Marti, S., Sawhney, N., Wheeler, S. (2000). Everywhere messaging. IBM Systems Journal, Volume 39, Numbers 3 & 4. http://www.research.ibm.com/journal/sj39-34.html

[180] Schmandt, C., Marti, S. (2005). Active Messenger: Email Filtering and Delivery in a Heterogeneous Network. Human-Computer Interaction Journal (HCI), Volume 20 (2005), Numbers 1 & 2.

[181] Schmidt, A., Aidoo, KA., Takaluoma, A., Tuomela, U., Van Laerhoven, K., Van de Velde, W. (1999). Advanced Interaction in Context. 1st International Symposium on Handheld and Ubiquitours computing (HUC99), Karlsruhe, Germany, Springer, pp 89-101. http://www.comp.lancs.ac.uk/~albrecht/pubs/pdf/schmidt_huc99_advanced_interaction_context.pdf

[182] Schmidt, A., Gellersen, H. W. (2001). Context-Aware Mobile Telephony. SIGGROUP Bulletin 22(1), pp 19-21. http://doi.acm.org/10.1145/500721.500726

[183] Schmidt, A., Takaluoma, A., Mäntyjärvi, J. (2000). Context-Aware Telephony over WAP. Personal Technologies 4(4), December 2000, pp 225-229. http://www.teco.uni-karlsruhe.de/~albrecht/publication/huc2k/context-call-draft.pdf

[184] Schneider, J., Kortuem, G., Preuitt, D., Fickas, S., Segall, Z. (2001) Auranet: Trust and Face-to-Face Interactions in a Wearable Community. Technical report. http://www.cs.uoregon.edu/research/wearables/Papers/auranet.pdf

[185] Sculley, J. (1987). Odyssey: Pepsi to Apple... A Journey of Adventure, Ideas, and the Future. New York, NY: Harper and Row.

[186] Sculley, J. (1989). The relationship between business and higher education: A perspective on the 21st century. Communications of the ACM 32, 9 (September), pp 1056-1061. http://doi.acm.org/10.1145/66451.66452

[187] Sekiguchi, D., Inami, M., Tachi. S. (2001). RobotPHONE: RUI for Interpersonal communication, Proceedings of CHI 2001, (Seattle, USA 2001), pp 277-278. http://doi.acm.org/10.1145/634067.634231

[188] Selker, T., Burleson, W. (2000). Context-aware design and interaction in computer systems. IBM Systems J. 39, 3&4, pp 880–891. http://researchweb.watson.ibm.com/journal/sj/393/part3/selker.html

[189] Severinson-Eklundh, K., Green, A., Hüttenrauch, H. (2003). Social and collaborative aspects of interaction with a service robot. Robotics and Autonomous Systems, Special Issue on Socially Interactive Robots, vol. 42, no. 3-4.
ftp://ftp.nada.kth.se/IPLab/TechReports/IPLab-208.pdf
http://vrai-group.epfl.ch/projects/iros2002/presentations/eklundh.pdf

[190] Shell, J., Selker, T., Vertegaal, R. (2003). Interacting with Groups of Computers. In Special Issue on Attentive User Interfaces, Communications of ACM 46(3), March 2003, pp 40-46.
http://doi.acm.org/10.1145/636772.636796

[191] Shell, J., Vertegaal, R, Skaburskis, A. (2003). EyePliances: Attention-Seeking Devices that Respond to Visual Attention. In Extended Abstracts of ACM CHI 2003 Conference on Human Factors in Computing Systems, 2003.
http://doi.acm.org/10.1145/765891.765981

[192] Singh, P. (2002). The public acquisition of commonsense knowledge. Proceedings of AAAI Spring Symposium on Acquiring (and Using) Linguistic (and World) Knowledge for Information Access. Palo Alto, CA: AAAI.
http://web.media.mit.edu/~push/AAAI2002-Spring.pdf

[193] Singh, P., Minsky, M. (2003). An architecture for combining ways to think. To appear in Proceedings of the International Conference on Knowledge Intensive Multi-Agent Systems. Cambridge, MA.
http://web.media.mit.edu/~push/WaysToThink.pdf

[194] Singh, P., Williams, W. (2003). LifeNet: A Propositional Model of Ordinary Human Activity. Submitted to the Workshop on Distributed and Collaborative Knowledge Capture (DC-KCAP) at K-CAP 2003.
http://web.media.mit.edu/~push/LifeNet.pdf

[195] Strommen, E. (1998). When the Interface is a Talking Dinosaur: Learning Across Media with ActiMates Barney. In Proceedings of the ACM CHI 98 Human Factors in Computing Systems Conference, pp 288-295.
http://doi.acm.org/10.1145/274644.274685

[196] Suzuki, N., Bartneck, C. (2003). Subtle Expressivity of Characters and Robots. In Proceedings of the CHI2003, Fort Lauderdale, USA.
http://doi.acm.org/10.1145/765891.766150

[197] Tang, J., Yankelovich, N., Begole, J., Van Kleek, M., Li, F., Bhalodia, J. (2001). ConNexus to Awarenex: Extending awareness to mobile users. In Proceedings of ACM CHI 2001, Seattle, Washington, March 31-April 5, pp 221-228.
http://doi.acm.org/10.1145/365024.365105

[198] Thomas, F., and O. Johnson. (1981). Disney Animation: The Illusion of Life. New York, Abbeville Press.

[199] Van de Velde, W. (1997). Co-Habited Mixed Reality. Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence (IJCAI-97), Aichi, Japan, August 23-29, 1997.

[200] Van de Velde, W. (1999). On the Self-Evaluation of a Wearable Assistant. In H. Gellersen (Ed.) Handheld and Ubiquitous Computing (HUC '99), Lecture Notes in Computer Science No. 1707, Springer-Verlag Heidelberg: 1999, pp 325-327.

http://www.springerlink.com/openurl.asp?genre=article&issn=0302-9743&volume=1707&spage=324

[201] Vertegaal R., Slagter R., Van der Veer, G., Nijholt, A. (2000). Why conversational agents should catch the eye. Proceedings ACM SIGCHI Conference CHI 2000, pp 257-258. http://doi.acm.org/10.1145/633292.633442

[202] Vertegaal, R., Dickie, C., Sohn, C. & Flickner, M. (2002). Designing Attentive Cell Phones Using Wearable EyeContact Sensors. Proceedings of CHI2002, pp 646-647. http://doi.acm.org/10.1145/506443.506526

[203] Vertegaal, R., Slagter, R., Van der Veer, G., Nijholt, A. (2001). Eye Gaze Patterns in Conversations: There Is More to Conversational Agents than Meets The Eyes. Proceedings of CHI2001, pp 301-308. http://doi.acm.org/10.1145/365024.365119

[204] Weiss, P. (2003). Minding Your Business: Humanizing gadgetry to tame the flood of information. Science News, May 3, 2003; Vol. 163, No. 18, p 279. http://www.sciencenews.org/20030503/bob8.asp

[205] Winston, P. H. (1992). Artificial intelligence (3rd ed.). Reading, Mass.: Addison-Wesley Pub. Co.

[206] Woodruff, A., Aoki, P. M. (2003a). How Push-to-Talk Makes Talk Less Pushy. Proc. ACM SIGGROUP Conference on Supporting Group Work (GROUP '03), Sanibel Island, FL, November 2003, pp 170-179. http://doi.acm.org/10.1145/958160.958187

[207] Woodruff, A., Aoki, P. M. (2003b). Media Affordances of a Mobile Push-to-Talk Communication Service. Technical report, February 2003. http://www2.parc.com/csl/projects/audiospaces/pdf/Woodruff-2003-PushToTalk.pdf

[208] Yonezawa, T., Clarkson, B., Yasumura, M., Mase,K. (2001). Context-aware Sensor-Doll as a Music Expression Device. In Extended Abstracts of CHI' 01, ACM Press, pp 307-308. http://doi.acm.org/10.1145/634067.634249

[209] Zlatev, J. (1999). The Epigenesis of Meaning in Human Beings and Possibly in Robots. Lund University Cognitive Studies, vol.79. http://www.lucs.lu.se/People/Jordan.Zlatev/Papers/Epigenesis.pdf

Version (b) of May 6th, 2005