Physical Embodiments for Mobile Communication Agents

Stefan Marti and Chris Schmandt
Speech Interface Group
MIT Media Lab, 20 Ames Street
Cambridge, MA 02139 USA

ABSTRACT

This paper describes a physically embodied and animated user interface to an interactive call handling agent, consisting of a small wireless animatronic device in the form of a squirrel, bunny, or parrot. A software tool creates movement primitives, composes these primitives into complex behaviors, and triggers these behaviors dynamically at state changes in the conversational agent's finite state machine. Gaze and gestural cues from the animatronics alert both the user and co-located third parties of incoming phone calls, and data suggests that such alerting is less intrusive than conventional telephones.

ACM Classification: H5.2. [Information interfaces and presentation]: User Interfaces: Interaction Styles.

General terms: Design**,** Human Factors

Keywords: Embodiment, robotic user interface, conversational agent, human style non-verbal cues, interruptions

INTRODUCTION

The telephone is a device we love to hate, but we *cannot live without it* [15]. Its interruption is important to our productivity at work and social and familial availability, yet we detest a distracting call breaking up a face-to-face conversation. Those around us, co-located conversation partners or strangers who happen to share the physical space, are also impacted. These third party eavesdroppers can become uncomfortable and annoyed by the interrupting call that has nothing to do with their ongoing activity, and may behave so as to assert their own physical presence [17]. The ubiquity of mobile phones guarantees that these interruptions impact most aspects of our daily lives.

There have been a number of conversational telephone managing agents, such as PhoneSlave [22] and the commercialized Wildfire™. These audio-based agents converse with a caller mainly to re-direct a telephone call. More recently, an effort has been made to address the possibility of adding the recipient's context to the agent's decision matrix [23]. We believe that recipient's *social* context should be incorporated into the agent's decision as well, but context aware systems are often brittle and inaccurate in their assessment of the user's context without additional human input. Since the user is part of a social setting that includes local others, the manifestation of the agent needs to behave in a socially appropriate way so that the interaction between user and agent does not become more disruptive than a phone call would be anyway. Because the embodiment of the agent has to blend into the reality of the user and his or her co-located conversation partners, it must be a physical entity that inhabits the same physical space as we do, not just a character on a screen [13].

In this paper, we describe and focus on the user interface of a software agent with the following features: The agent detects conversations to determine social groupings, tries to evaluate the importance of incoming communication, invites input from the local others, while also consulting memory of previous interactions stored in the location. It interacts with the caller, the callee, and co-located others, and may, e.g., answer the call, offer some information, and suggest leaving a voice instant message instead of disrupting an ongoing conversation. In turn, it may allow the caller to wait while the recipient hears this message and chooses how or whether to respond.

The software agent that we present is developed on a desktop computer, but controls its embodiment wirelessly (remote brain approach). We anticipate that eventually either the agent will run directly on a cellphone and control the embodiment via a low-range wireless link, or that the phone itself becomes part of the embodiment. Our most recent prototype has a Bluetooth transceiver for both audio and data link, and the only bottleneck is the computational power of available Bluetooth cellphones that may not yet suffice to run all processes of our agent system.

APPROACH

This paper focuses on 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 [21].

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 [29] by sending out readable social cues that convey its internal state. We do not imply that our agent's software mimics mental cognitive processes. However, it is designed to express itself

UIST'05, October 23–27, 2005, Seattle, Washington, USA.
Copyright 2005 ACM 1-59593-023-X/05/0010…$5.00.

with human-style non-verbal cues such as gaze and gestures to generate certain effects and experiences with the user. We are convinced that human-style social cues can improve the affordances and usability of an agent system.

Key to this work is giving a conversational agent physical presence, through interactive “stuffed animals” of different shapes and sizes, remotely controlled by a computer (Figure 1). 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 could be ideal for unobtrusive interruption by 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 [9]. Furthermore, these cues also 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.

Our 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” [6]. 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 [18], the agent may choose to interrupt the user only when there is no speech activity.

We use zoomorphic embodiments combined with anthropomorphic behaviors (gaze, gestures). Although this combination partially violates the ‘life-likeness’ of our crea-

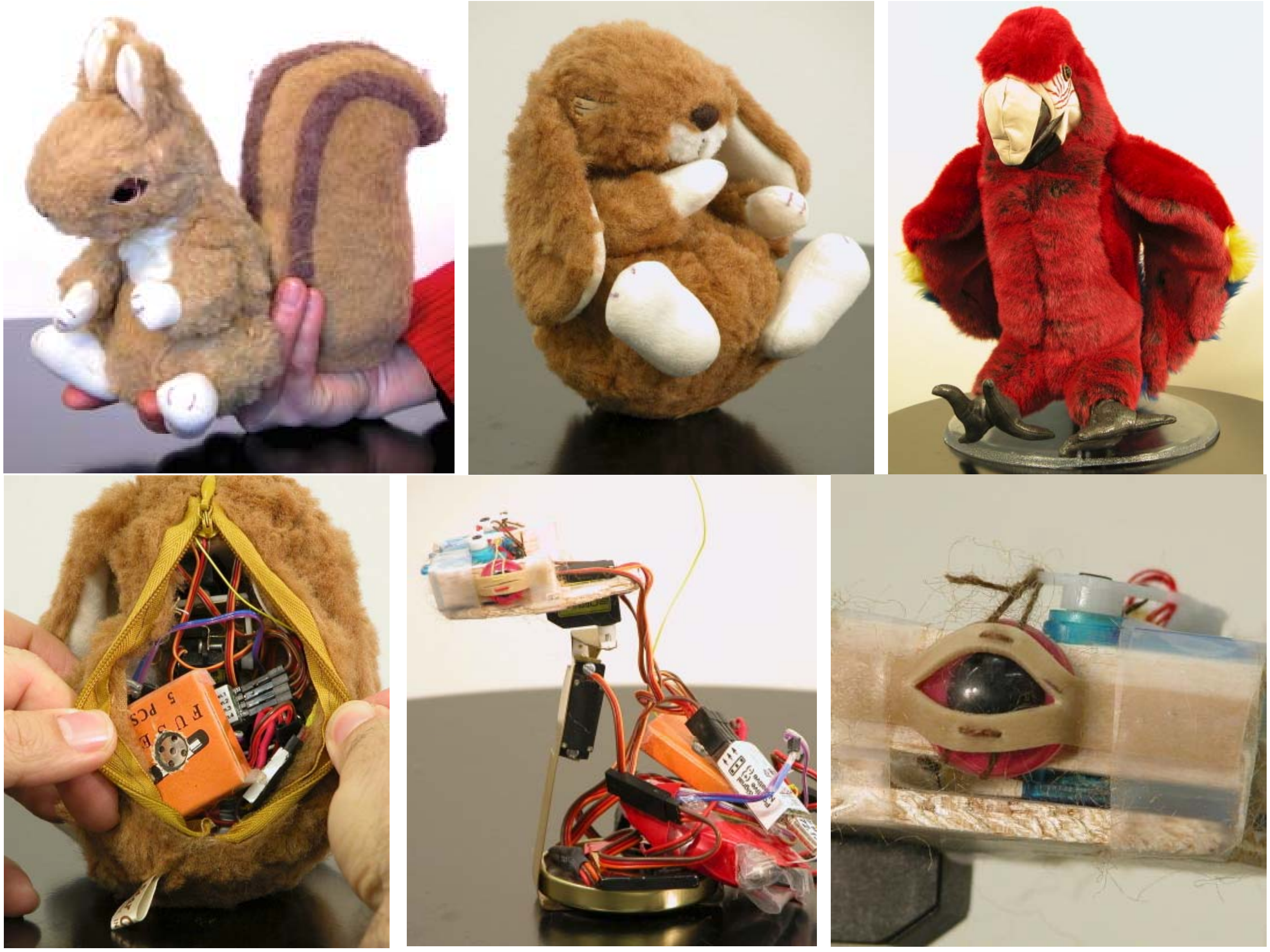


Figure 1: Top row, left to right: squirrel (11cm tall), bunny (10cm tall); parrot (38cm tall). Bottom row: bunny with open back, bunny skeleton, eye and lid mechanics.

tions, it also allows us 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.

We recognize that proposing animatronic stuffed animals as mobile telephone sets is somewhat controversial. We started this work with some caution, noting an affinity for personalized and "soft" telephone covers and sets in Japan and Korea. As we built the three versions documented in this paper, we found that people exhibited a strong curiosity and affinity for our prototypes. They often evoke stories about people's experiences with their pets. As Greenberg [7] notes, physical and embodied computer interfaces are in their infancy, and we encourage thinking about them with an open mind while exploring a diversity of forms, as his work does. In particular, with mobile telephony we are already dealing with devices with high emotional impact, so it is perhaps encouraging that a radical design such as ours also evokes strong responses.

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' [4]. The more the system can interact with its environment, the more it is embodied.

In our 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.

Although our current embodiments are all based on animals (bunny, squirrel, and parrot), their respective morphologies are diverse enough so that their appearances create different expectations (and preferences, as our user study shows). These expectations influence the behaviors that the user might want to see from the animatronics. Due to our 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 we foresee that users will have strong individual preferences for their personal animatronics.

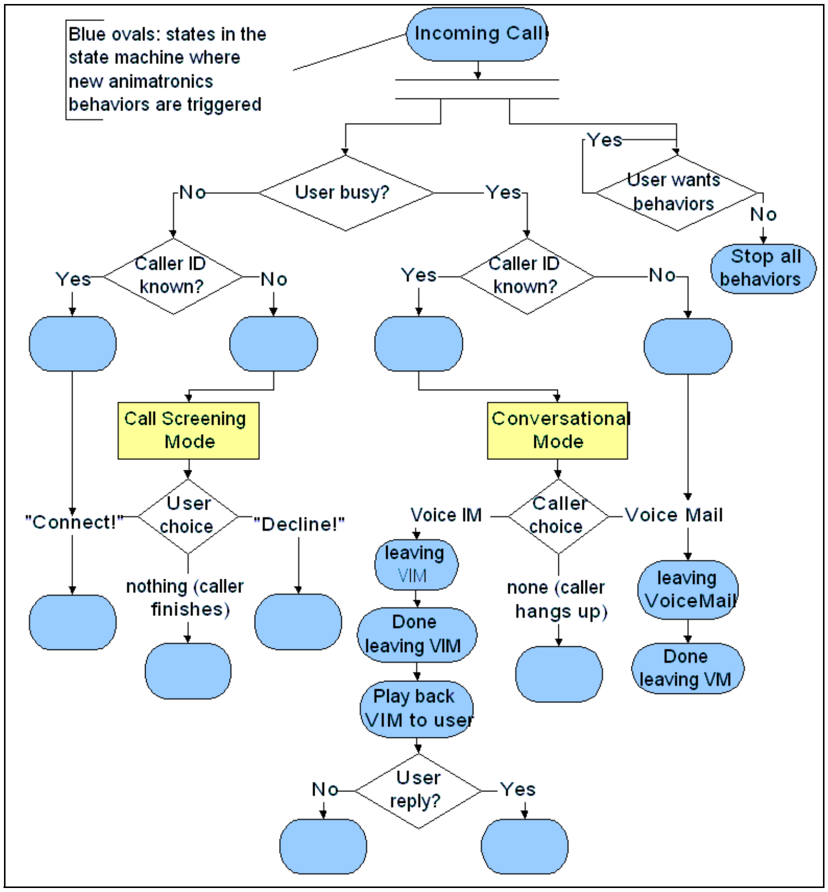


Figure 2: Basic conversational call tree

SYSTEM

Our current system consists of the following elements:

- Small animatronic devices (squirrel, bunny, parrot); carried or worn by user
- Computer-remote control for the animatronics: animatronics control server and wireless link (Bluetooth transceiver)
- Software tool for building libraries of atomic behaviors, or primitives, and for composing those primitives into gestures and behaviors
- Dual conversational agent that converses with both caller and user; implemented as a finite state machine
- Means of scripting gestures and behaviors to conversational agent states

Dual conversational agent

Our conversational agent can converse with both the caller and the user, at the same time. The interaction on the caller side is audio only. The agent can use both non-verbal cues and spoken language in its interaction with the user. The conversational agent is implemented as a finite state machine. It follows a decision tree with branches that depend on external data and sensors, as well as caller, user, and co-located people's choices, which are detected via speech recognition and button presses. The following are the main factors influencing state changes:

- Distinction between important and unimportant callers by matching caller ID against a list of preferred 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, allowing others in the conversation input as whether the call should go through [18], and knowing how other people in this location have responded to incoming calls (not discussed in this paper)

Figure 2 shows an overview of the basic call tree, which is a subset of the current functionality. For example, if the user is busy (inferred by detecting that she is in a conversation), and receives a call from a recognized number, the agent answers the phone and offers to take a voice mail or voice instant message, in which case the caller speaks a message and waits.

Animatronics, computer-remote control

The animatronics are stuffed animals enhanced with custom made skeletons, actuators, transceivers, batteries, etc.

Our 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, 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.

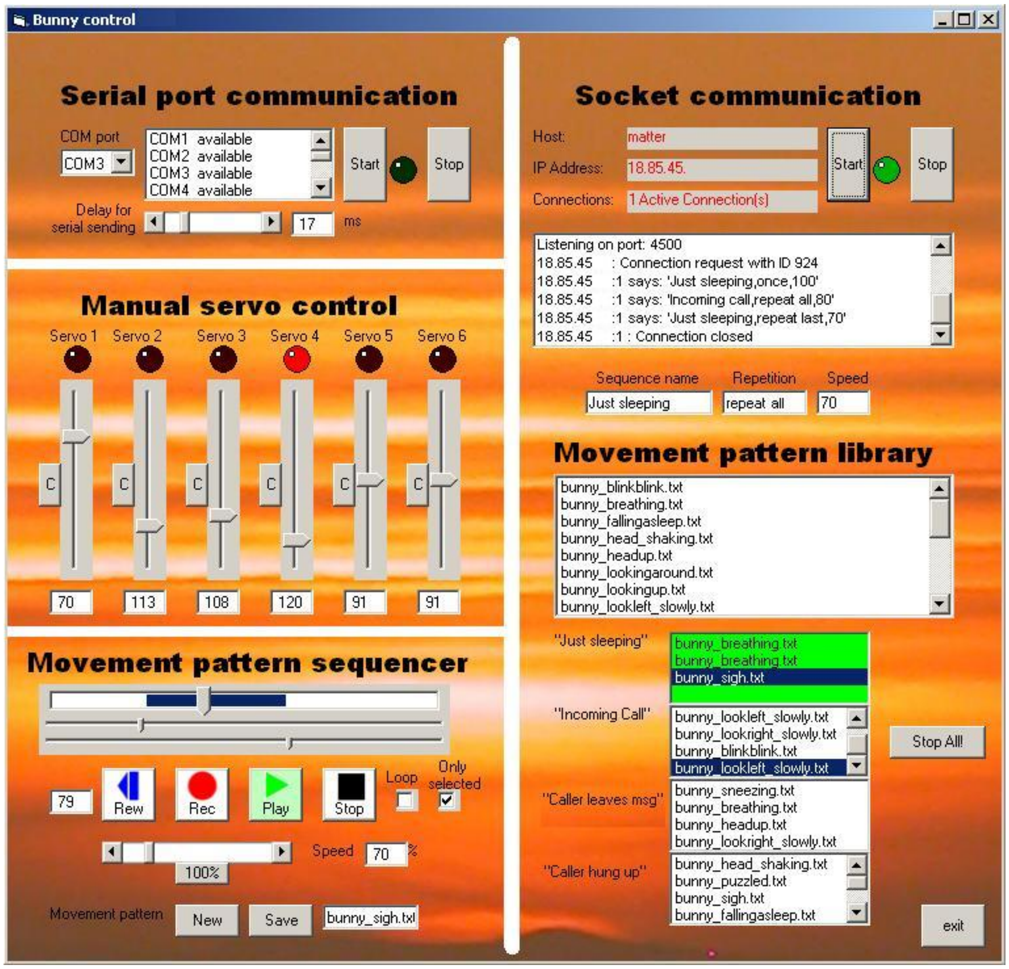
In order to create a realistic eye opening and closing expression, both bunny and squirrel are able to move both upper and lower lids, using small rubber bands as lids that are pulled back simultaneously by a micro servo via thin threads. All actuators are independent channels that are fully proportional with a resolution of 100 steps from one extreme to the other.

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

Although in the future, the animatronics may be controlled directly by the user's cellphone, or the animatronics contains the cellphone, our current animatronics prototypes are implemented with a 'remote brain' approach: they are computer-remote controlled, but completely wireless and self-contained devices. We have built three generations of embodiments that differ in their capabilities:

1. Parrot: simplex data link, no audio
2. Bunny: simplex data link; half-duplex audio
3. Squirrel: full duplex audio and data link

In order to control the parrot and the bunny, we modified radio control gear that is used by hobbyists to control airplanes and boats. This channel is simplex, with a range of up to 100 meters indoors. Our animatronics control software sends out serial signals over RS232 to a "glue" board containing a microcontroller that generates a proprietary pulse width modulation signal, which is fed into a 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


Bunny control

Serial port communication

COM port: COM3

COM1 available
COM2 available
COM3 available
COM4 available

Delay for serial sending: 17 ms

Manual servo control

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

Movement pattern sequencer

79

Loop □ Only selected ☑

Speed 70 %

100%

Movement pattern: bunny_sigh.txt

Socket communication

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	Repetition	Speed
Just sleeping	repeat all	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 3: Screenshot of animatronics sequencer and server

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 and helicopters.

Our 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. 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.

Our most advanced embodiment, the squirrel, sports a fully digital link for both audio and data. On the desktop computer side, we use a Bluetooth class 1 transceiver with modified antenna to achieve a range of 40 meters indoors. On the embodiment side, we use a Bluetooth class 1 module with a ceramic antenna. 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 us 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 enables us to send back sensor data from the embodiment to the animatronics control software.

Behavior primitives, composite behaviors, and synchronizing agent states with composite behaviors

All our embodiments are controlled remotely by our animatronics server and sequencer (Figure 3). 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. In the future, we anticipate that the software with the hub functionality will run directly on the user's phone, whereas the authoring tool may remain on a desktop. The current animatronics server incorporates the following functions:

- Record and modify behavior primitives in loops
- Compose primitives into behavior sequences
- Map complex behavior sequences to the conversational agent's state changes

Creating behavior primitives

At the core of the animatronics control software is the **Manual Servo Control**, which allows the character designer to manipulate each DOF separately via sliders. The manipulation of DOFs is used in the **Movement Pattern Sequencer**, 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 manipulates 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 'overwriting' the change with a new one during the next loop. This recording metaphor is similar to the 'audio dubbing' method used in movie making, where the actor watches short scenes in a loop, and can keep recording and adjusting the dubs until satisfaction.

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. 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 she just did, as well as the animatronics following the slider movements.

A 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. A primitive can also be pruned at its beginning and end via horizontal sliders. Once a primitive is created and modified to the designer's satisfaction, it can be named and stored in the **Movement Pattern Library**, and recalled at any time.

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 for 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

Mapping behaviors to agent states

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.

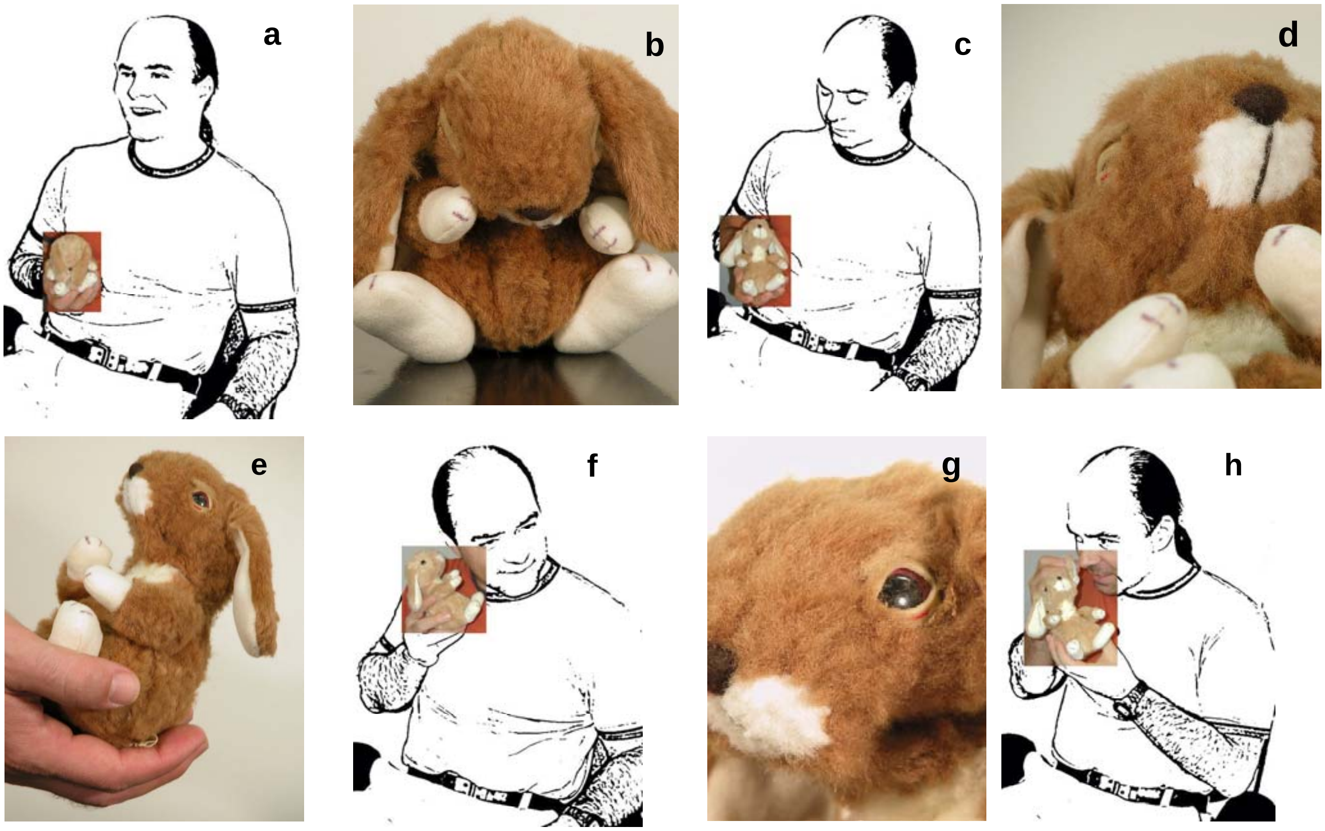


Figure 4: Top row, left to right: bunny sleeping, waking up, listening to caller
Bottom row, left to right: trying to get attention with gaze, whispering to user, being attentive, listening to user

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 activated in real-time, depending on the agent-caller interaction. Therefore, the behaviors vary not only by user and caller actions, but also on factors such as the length of a voice instant message.

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.

Short video clips of behaviors can be found at:
http://web.media.mit.edu/~stefanm/phd/videos/

Interaction example

In the example below we show the relationship between state transitions, the animatronics' behavior we intend to express, 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 (Figure 4a). Every now and then it sighs (moves head twice up and down, 10% of the actuator travel) in order to let its owner know that everything 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 (Figure 4b). The agent 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%, Figure 4c).

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 (Figure 4d). 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 (Figure 4e). The bunny whispers in his ear and tells him who is on the phone, then plays back the short message it took from Clara (Figure 4f). The animatronics is now fully awake and attentive (eyes completely open, head straight) (Figure 4g). 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 often) (Figure 4h). 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.*

EVALUATION

For handling phone calls, we believe that a physical embodiment facilitates the mental separation between talking to remote others and co-located people. But these claims are hard to validate. The first requires extended use of robust hardware, and it is only in our all-digital most recent embodiment that we have begun to approach portability with acceptable range. Due to the novelty of "talking to a stuffed animal," the second 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.

Still, we can 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 [17], we decided to interview participants while staging interruptions using both conventional and animatronic telephones. Participants' reactions were examined by observing their videotaped behavior, evaluating Semantic Differential questionnaires, and comments from semi-structured post-exposure interviews.

Experimental procedure

We tested 10 participants (age 25 to 55; 4 male, 6 female). They were administrative or support staff from our building who had little or no previous contact with the project. Each trial 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 by Omron, etc.), and the numerous stuffed animals that routinely adorn the interviewers computer monitors. For the first five to ten minutes, 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. A Semantic Differential [20] 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 our 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.

Results and Discussion

Quantitative results

Our null hypothesis was that attitudes towards interruption would be independent of whether interruption was by a traditional ringing telephone or a moving animatronic device. Our 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 1, 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 1: 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**. (A classic EPA analysis was not attempted because of too low N.) There were no statistically significant differences due to gender or recency—i.e., the most recently experienced interruption was not more annoying.

The 17 scales that were used are specified by the Semantic Differential protocol, and are constructed to measure how conditions evoke different feelings towards concepts, devices, or interfaces; they do not directly measure preference. Semantic Differentials 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 to chose between ringing phone and animatronics interruption, may not consistently prefer one over the other, their *affective* attitudes towards the two choices differ significantly and consistently. The results clearly show a strong difference in reactions to the two conditions; to understand the implications of these differences, we resorted to qualitative techniques.

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 objectionable [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 than 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.”

Our 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 our small sample size, reactions to the phone and squirrel conditions were so different that we quickly obtained statistically significant results, and for that reason did not run more subjects. Since it is based on a large number of dimensions, a complete Semantic Differential would have required many more subjects. Our subjects' comments and our analysis of their reactions both by the interviewer and later on videotape were rich; the quotes above are representative, but only a small fraction of the total.

We did, however, find some limitations or reservations by our subjects, mostly around the particular animal forms chosen, and clearly some sensitivity to the sounds made by some of our de-sensitizing props (which were active mostly while subjects read consent forms). For example, referring to our 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.”

RELATED WORK

Physical embodiments as user interfaces have been studied and applied in a variety of contexts. A *Robotic User Interface* (e.g., Bartneck et al.) [1] is the paradigm where robots are used as an interface between the physical world and information world. As an example, Kuzuoka et al. [14] and Greenberg et al. [8] suggest digital but physical surrogates in an office environment. They are digital representations of people (avatars), something our agent does not intend to be. Jabarin et al. [11] 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 we share.

Our work is also in the tradition of *Socially Intelligent Agents,* which is based on Reeves and Nass’ findings of “computers as social actors.” [21] Kismet (Breazeal [2]) is a prominent example of a socially intelligent robot. Although it has not the same function as our agent, it demonstrates the importance of socially strong nonverbal cues to grab attention, show interest, etc. Breazeal et al. [3] 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. Okada et al. [19] look at the important social bonding between artificial autonomous creatures (such as cyberpets) and humans, especially its conversational aspects.

Suzuki et al. [25] initiated work under the label of *Subtle Expressivity for Characters and Robots*, an idea that resonates significantly with Hansson et al.’s [9] work on subtle but public alerts in communication. Two relevant papers in this context are Liu et al. [16] and Isbister [10].

Our system is also related to *Intelligent Interface Agent* research. One of the first embodiments of an agent as a bird was probably the COMRIS parrot (Co-Habited Mixed Reality Information Spaces) by Van de Velde [26] and De Haan [5]. 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 by which it influences its wearer’s behavior. Van de Velde [27] also looks at the effectiveness of such wearables as an advice giver. Kaminsky et al. [12] describe 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 related, since the authors use robotic 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. [28] describe a sensor doll for musical expression that is capable of multi-modal interaction with the user. The doll is an embodied agent that displays 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 [24], which seems to solve a similar problem as our 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 our system. We regard our animatronics as entities independent from caller and user, whereas RobotPHONE does not make that claim.

CONCLUSION

In this paper we have discussed the use of wireless animatronic stuffed animals as user interface embodiments of communication agents. We described three generations of wireless devices, culminating in a Bluetooth version supporting full duplex audio. A software GUI tool allows character designers to create behavior primitives by manipulating the animatronics' degrees of freedom via sliders, and to composite those primitives into complex behaviors. A small user evaluation suggests that such animatronics evoke significantly different reactions than ordinary telephones and are seen as less invasive by others present when we receive phone calls.

ACKNLOWLEDGEMENTS

We would like to thank Matthew Hofmann, with support from the Shih-Ming Shih (1979) Fund, for his work programming the conversational agent, Kimiko Ryokai, all the participants of our study, and Ben for telling us that it is okay for boys to play with stuffed animals.

REFERENCES

1. Bartneck, C., Okada, M. (2001). Robotic User Interfaces. Proceedings of the Human and Computer Conference (HC-2001), Aizu, pp 130-140.
2. Breazeal, C. L. (2002). Designing sociable robots. Cambridge, MA: MIT Press.
3. Breazeal, C., Fitzpatrick, P. (2000). That Certain Look: Social Amplification of Animate Vision. AAAI Fall Symposium on Socially Intelligent Agents: The Human in the Loop, Technical Report FS-00-04, pp 18-23.
4. 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.
5. 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, pp 69-79.
6. Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2002). A Survey of Socially Interactive Robots: Concepts, Design, and Applications. Technical report CMU-RI-TR-02-29, Robotics Institute, CMU.
7. Greenberg, S. (2004). Collaborative Physical User Interfaces. Report 2004-740-05, Department of Computer Science, University of Calgary, Alberta, Canada.
8. Greenberg, S. and Kuzuoka, H. (2000). Using Digital but Physical Surrogates to Mediate Awareness, Communication and Privacy in Media Spaces. Personal Technologies, 4(1), pp 182-198.
9. Hansson, R., Ljungstrand, P., Redström, J. (2001). Subtle and Public Notification Cues for Mobile Devices. Proceedings of UbiComp 2001, pp 240-246.
10. 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.
11. Jabarin, B., Wu, J., Vertegaal, R., Grigorov, L. (2003). Establishing Remote Conversations Through Eye Contact With Physical Awareness Proxies. Extended Abstracts of CHI 2003, pp 948-949.
12. Kaminsky, M., Dourish, P., Edwards, K. LaMarca, A., Salisbury, M., Smith, I. (1999). SWEETPEA: Software tools for programmable embodied agents. Proceedings of CHI'99, pp 144-151.
13. Kidd, C., Breazeal, C. (2004). Effect of a Robot on Engagement and User Perceptions. Proceedings of International Conference on Intelligent Robots and Systems (IROS04), Sendai, Japan, pp 3559-3564.
14. Kuzuoka, H., Greenberg, S. (1999). Mediating Awareness and Communication through Digital but Physical Surrogates. Extended Abstracts of CHI'99, pp 11-12.
15. Lemelson-MIT Invention Index study, 8th annual, January 21, 2004, online at: http://web.mit.edu/invent/n-pressreleases/n-press-04index.html
16. Liu, K.K., Picard, R.W. (2003). Subtle Expressivity in a Robotic Computer. CHI 2003 Workshop on Subtle Expressivity for Characters and Robots.
17. 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.
18. Marti, S., Schmandt, C. (2005). Giving the Caller the Finger: Collaborative Responsibility for Cellphone Interruptions. Extended Abstracts of CHI 2005.
19. 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.
20. Osgood, E.C., Suci, G.J., & Tannenbaum, P.H. (1957). The measurement of meaning. Urbana: University of Illinois Press.
21. Reeves, B., Nass, C. I. (1996). The media equation: how people treat computers, televisions, and new media like real people and places. Stanford, CA: CSLI Publications/Cambridge University Press.
22. Schmandt, C., Arons, B. (1984). A Conversational Telephone Messaging System. IEEE Transactions on Consumer Electronics CE-30, 3 (Aug 1984), pp 21-24.
23. Schmidt, A., Aidoo, KA., Takaluoma, A., Tuomela, U., Van Laerhoven, K., Van de Velde, W. (1999). Advanced Interaction in Context. In H. Gellersen (Ed.) Handheld and Ubiquitous Computing (HUC '99), Lecture Notes in Computer Science No. 1707, Springer-Verlag Heidelberg: 1999, pp 89-101.
24. Sekiguchi, D., Inami, M., Tachi, S. (2001). RobotPHONE: RUI for Interpersonal Communication. Extended Abstracts of CHI 2001, pp 277-278.
25. Suzuki, N., Bartneck, C. (2003). Subtle Expressivity of Characters and Robots. Extended Abstracts of CHI 2003, pp 1064-1065.
26. Van de Velde, W. (1997). Co-Habited Mixed Reality. Fifteenth International Joint Conference on Artificial Intelligence (IJCAI-97), Aichi, Japan.
27. 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.
28. Yonezawa, T., Clarkson, B., Yasumura, M., Mase,K. (2001). Context-aware Sensor-Doll as a Music Expression Device. Extended Abstracts CHI'01, pp 307-308.
29. Zlatev, J. (2001). The Epigenesis of Meaning in Human Beings and Possibly in Robots. Minds and Machines 11(2), pp 155-195.