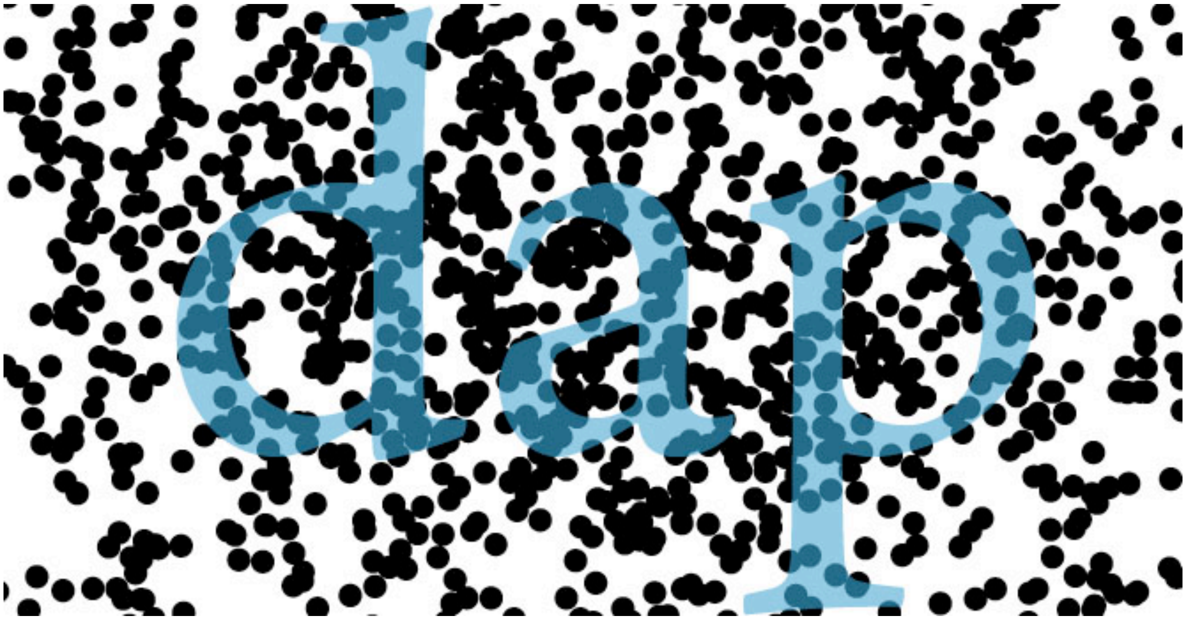


Devices that Alter Perception

DAP2008

A workshop organized in conjunction with UbiComp 2008

Seoul, South Korea, September 21st, 2008

Table of Contents

Preface ········ iii

Carson Reynolds, Alvaro Cassinelli, Anthony Dunne, and James Auger

Feel the Force: Using Tactile Technologies to Investigate the Extended Mind ········ 1

Jon Bird, Simon Holland, Paul Marshall, Yvonne Rogers and Andy Clark

SocialSense: A System For Social Environment Awareness ········ 5

Robert S. Brewer, Samuel R. H. Joseph, Guanghong Yang, Neil Scott, and Daniel Suthers

Boxed Ego ········ 9

Alvaro Cassinelli and Masatoshi Ishikawa

Fear tuners – Prostheses for instincts ········ 13

Susanna Hertrich

Gesture recognition as ubiquitous input for mobile phones ········ 16

Gerrit Niezen and Gerhard P. Hancke

Creation of Sympathetic Media Content ········ 20

Stéphane Perrin, Giuseppe Riva, and Alvaro Cassinelli

Aural Antennae ········ 25

Carson Reynolds, Alvaro Cassinelli, and Masatoshi Ishikawa

Learn Traffic State Based on Cooperative Localization ········ 29

Zhuo Sun, Yiqiang Chen, and Juan Qi

Spatial coverage vs. sensorial fidelity in VR ········ 33

Alexis Zerroug, Alvaro Cassinelli, and Masatoshi Ishikawa

Devices that Alter Perception (DAP 2008)

Carson Reynolds **Alvaro Cassinelli**
University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8658, Japan
{carson, alvaro}@k2.t.u-tokyo.ac.jp

Anthony Dunne **James Auger**
Royal College of Art
Kensington Gore, London SW7 2EU, UK
anthony.dunne@rca.ac.uk info@augerment.com

PREFACE

Devices that Alter Perception is a new workshop that aims to instigate development and critique of systems that focus on the human percepts. Sensors, actuators, implants, wearable computers, and neural interfaces can do more than simply observe our bodies; these devices can alter and manipulate our perceptions. The goals of the workshop are to: (1) better understand the process of perception (2) aid those developing devices by sharing designs (3) debate of ethical and social issues that are unique to devices that operate below or upon awareness.

Accepted position papers are presented in 10-minute oral presentations or demonstrations followed by 5-minute question and answers sessions. Additionally, the position papers are uploaded to a special discussion site (http://dap.reddit.com) for commentary as well as voting. The paper receiving the highest score–as determined by open, public voting–will be awarded a best paper prize.

ACKNOWLEDGEMENTS

We would like to thank the Sunny Consolvo, Hyung Kyu Song, Jongwon Kim, Timothy Sohn and the UbiComp organizers for the tremendous amount of help provided. We would also like the thank Masatoshi Ishikawa for providing a supportive environment for the development of this research theme.

ON LINE MATERIALS

The call for papers, information for attendees, and accepted submissions are hosted at:

http://www.k2.t.u-tokyo.ac.jp/perception/dap2008/

SPONSORS

This workshop is jointly sponsored by the University of Tokyo's Meta-Perception Research Group and the Royal College of Art's Design Interactions Department.

Feel the Force: Using Tactile Technologies to Investigate the Extended Mind

Jon Bird, Simon Holland, Paul Marshall,
Yvonne Rogers
The Open University
Pervasive Interaction Lab, Computing
Department, Milton Keynes, MK7 6AA, UK
{J.Bird, S.Holland, P.Marshall,
Y.Rogers}@open.ac.uk
+44(0)1908 274066

Andy Clark
University of Edinburgh
Department of Philosophy, George Square,
Edinburgh EH8 9JX, UK
Andy.Clark@ed.ac.uk
+44(0)131 650 3659

ABSTRACT

We describe the motivations behind the E-Sense project which will investigate augmented perception by building a range of novel tactile interfaces. As well as exploring the practical utility of these systems for real world tasks, we are particularly interested in the following question: how can we design tactile interfaces to mediate novel sensory information so that the user experiences the technology as an extension of themselves?

Author Keywords

E-Sense, extended mind, transparent technologies, tactile interface

ACM Classification Keywords

B.4.2 Input/Output Devices, H5.m. Information interfaces and presentation, K.4.1.c Ethics

INTRODUCTION

Recent work in philosophy and cognitive science has introduced the idea of the extended mind (for example, [5]), a view of the human cognitive system as a plastic hybrid of biological and non-biological components, including external representations and technologies. This perspective has profound implications for our notion of what it means to be human, pointing to the potential to change thought and action by integrating new technologies and information sources.

Research into augmented perception[1] has established that a variety of sensory information can be mediated through tactile interfaces in a way that is understandable to users and can guide their actions. For example, in the pioneering work of Bach-y-Rita and co-workers on sensory substitution [1], blind participants have visual information from a camera represented to them in the form of the activation of an array of tactile actuators placed on their back, thighs or tongues. With practice, participants are able to use this tactile information to make perceptual judgements and co-ordinate action, for example batting a ball that is rolling off a table. Interestingly, as participants learn to use the tactile stimulation their perception of it changes: sensing the percept in space rather than on their skin. The interface becomes transparent in use, or 'ready-at-hand' to use Heidegger's phrase [6] - that is, the user experiences the technology as though it were an extension of themselves.

Neuroscience experiments have established that tool use can cause structural changes in the brain: the receptive fields of some neurons expand and incorporate the tool into the 'body schema' [12]. Significantly, the neuronal changes only occur when the tactile information is used to guide action, a finding that provides support for O'Regan and Noë's [13] characterisation of perception as primarily involving the mapping of sensorimotor contingencies: systematic relationships between action and sensory input. These perceptual mappings can be surprisingly plastic. Early work by Stratton [19] and Kohler [10] established that humans can adapt to radical disruptions of the relationship between sensors and actuators, for example, inverting glasses turning the visual field upside down. Of particular relevance to our project, Ramachandran and Blakeslee describe how the perceptual system can be tricked into producing the experience of having a two foot nose or experiencing tactile sensation in a table [15].

However, despite extensive citations in the literature, there is still substantial uncertainty concerning the nature of these augmenting sensory experiences. Given the remarkable capacity of people to adapt to changes in existing sensorimotor mappings and to incorporate novel sensory modalities, under what conditions does a mediating technology *not* become transparent? Does sensory extension support a 'sensorimotor contingencies' model of perceptual experience? If it does, what can we learn about the form of sensorimotor contingency mappings that remain 'opaque' and do not become

---

[1] 'Augmented perception' encompasses both 'sensory extension' and 'sensory substitution', and is where technology provides access to environmental energy not available to a person's biological perceptual system (for example, IR or ultrasound). In the substitution case this is because of perceptual impairment, for example, an individual is blind or deaf.

incorporated into the body; if it does not, which models better explain the perceptual experience of sensory extension? Are the mappings between action and augmenting sensory input as plastic as those coordinating biological senses and motor systems? In the interdisciplinary E-Sense project we believe that by creating a wide array of tactile interfaces and monitoring both their use and the user experiences on an ongoing, day to day level, we will gain important insights into these questions.

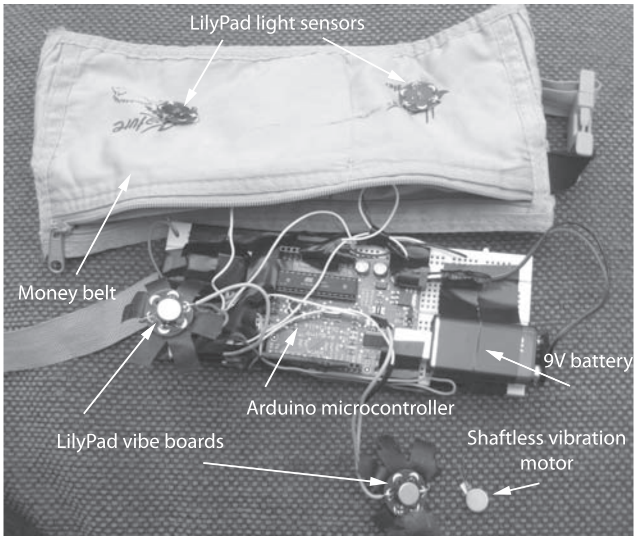


Figure 1. A rapid prototype built to test the suitability of Arduino LilyPad vibe boards for tactile sensory extension interfaces. If light levels go above a hard-wired threshold value, then each of the sensors switches on one of the vibe boards. The diameter of the shaftless vibration motor is 20mm. The LilyPad vibe boards consist of one of these motors mounted on a printed circuit board that enables users to connect them to a microcontroller using conductive thread and incorporate them into clothing.

METHODOLOGY

In our interdisciplinary approach conceptual philosophical analysis feeds into the design of the sensory augmentation systems and user studies will reciprocally feed back into philosophy. One concrete goal is to build useful sensory extension tools; another, more nebulous, goal is to generate novel insights into the extended mind. Our project is extremely open-ended as relatively little is known about the design issues related to tactile systems or about the conditions under which such technologies become transparent in use. Consequently, we believe a productive approach is to combine concepts and approaches from very different disciplines - psychology, philosophy and computer science. We are very aware of the potential pitfalls, as well as the benefits, that can result from interdisciplinary collaboration [18].

Rapid Prototyping Approach

We believe that a good way to develop and refine our conceptual thinking about the extended mind and sensory augmentation is to embody our ideas in physical artefacts and test them in the real world. This approach has been successful in the past, particularly in open-ended exploratory projects [3,4]. We want to complete as many iterations of the build-test-reflect cycle as possible during the project and so we are adopting a rapid prototyping approach to constructing sensory extension interfaces. We are using open source technologies such as the Arduino electronics prototyping platform [2] and the Processing programming language and environment [14] because with these tools we can quickly connect cheap, off-the-shelf components and build working prototypes. See Figure 1 for a prototype that was built in a few hours to test whether Arduino LilyPad vibe boards [11] were suitable actuators for a wearable tactile system. Constructing this prototype confirmed that these cheap shaftless motors do provide a clearly perceptible signal through clothing and also highlighted the advantage of building a system where the mapping between sensors and vibration motors is easily configurable.

The building blocks of our tactile interfaces will be reconfigurable modules, each of which will consist of up to 16 shaftless coin-type vibration motors (See Figure 1) - this is the maximum number that can be driven using Pulse Width Modulation (PWM) by a Texas Instruments TLC5940 chip. Modules can be daisy chained and driven by a single Arduino microcontroller. The motors will attach to garments using velcro so that their spatial arrangement can be changed quickly. The modules can mediate between behaviour and different environment energies simply by changing the sensors that are connected to the microcontroller. The mapping between the sensors and the vibration motors can be configured in software, as can interactions between the sensors (for example, we could implement lateral inhibition). This flexibility will allow us to rapidly configure different mappings between sensorimotor contingencies and explore the conditions under which the interface becomes transparent or remains opaque.

Evaluation

We plan to carry out the evaluations using a qualitative case study approach with a small number of participants. On going interviews and informal tests of performance will be conducted to investigate participants' phenomenal experience of using the technologies and to explore whether performance benefits might result. Findings from the empirical studies will be used to inform theoretical models as well as develop predictions about particular sensory extension systems.

EMPIRICAL STUDIES

We plan to build and test the three sensory extension systems summarised in Table 1 which details:

- where the tactile interface will be placed on a user's body
- the number of tactile modules and vibration motors
- the type of sensors connected to the system
- the motor actions that are mediated by the tactile interface - what is the system for?
- the initial mapping between the sensors and each tactile module

Prototype	Location of tactile interface	No. of tactile sensor modules and sensors	Sensor contingency	Motor contingency	Initial mapping
Tactile Car Seat	Back	1 (6)	Ultrasound	Sense close targets	topographic
Feel the Force	Waist	1 (8)	Virtual	Localize target	topographic
Exploring Harmony Space	Back	3 (48)	Pitch	Harmonic improvisation	topographic

Table 1. A comparison of the three prototype devices that we are planning on building with our configurable tactile interface

Tactile Car Seat

We propose to design a car seat that will provide the driver with a direct perceptual representation of objects in close proximity to the vehicle. We will use an array of 6 vibration motors driven by the activation of 6 ultrasonic sensors positioned on each side of the car at the front, middle and rear. The intensity of vibration will correspond to the proximity of objects to the associated sensor. We predict that with practice this information might improve drivers' situational awareness and increase vehicle safety. This is an important goal: approximately 50000 reports on road accident injuries or fatalities in the UK in 2005 listed failure to look properly as a contributing factor to the accident and approximately 1500 listed failure to see due to the vehicle blind spot [16].

The idea of using tactile representations of information in a car is not a new one. Ho, Tan and Spence [7], for example, describe how vibrotactile warning signals can be used to alert drivers to dangers on the road. However, these systems are designed to be attention grabbing and present information only at critical moments. We predict that presenting tactile information continuously through the car seat might increase the driver's feeling of connection to the car. In certain situations this could be advantageous, for example, enhancing a driver's ability to judge whether the car might fit into a tight parking space.

We will test the prototype interface using two 'quick and dirty' evaluation methods, neither of which will require a person to drive a real car. This is to avoid the heavy development overheads associated with designing for a real vehicle or complex high-end driving simulator. Firstly we will use the tactile interface to play 'blind man's buff' games where a blindfolded user seated in the lab has to detect the approach of people; and secondly, we will employ a Wizard-of-Oz approach linking movement in an off-the-shelf PC driving simulator with activation of the vibration motor module. While obviously very different from driving a real sensor augmented vehicle, these evaluation methods will enable us to rapidly gauge the potential of this interface to guide action and under what conditions it becomes transparent.

Feel the Force

This playful empirical study is inspired by the scene in Star Wars Episode IV: A New Hope where Luke Skywalker is getting his first training in the Force on the Millennium Falcon. He is wearing a helmet with an opaque visor that prevents him from seeing a flying robot that moves around him and occasionally zaps him with an electric shock. He has to 'feel the Force' in order to sense the position of the robot and block its zap with his light sabre.

Each user will wear a cummerbund containing 8 equally spaced vibration motors (45 degree separation). The user's 'light sabre' will consist of a Wii nunchuk connected to an Arduino microcontroller. Users will start in a 'registration' position and then the system will track their movements using the 3 axis accelerometer in the nunchuk. The aim of the game is to move the nunchuk so that it blocks zaps from a virtual robot. Its movement will be indicated by changes in activation across the array of vibration motors. A zap occurs when the robot gets closer, indicated by an increase in vibration intensity. If a user responds to this increase by moving the nunchuk to the correct position then they will get force feedback from a vibration motor attached to the nunchuk, indicating that they have blocked the zap; if they move to the wrong position then a number of vibration motors in the cummerbund will vibrate indicating they have been 'hit'.

We will measure how long it takes users to become proficient in blocking zaps. If combined with interviews, then one might be able to determine whether transparency, if achieved, is signalled by performance level. We can map any of the locations in virtual zap space to the vibration motors and explore how different mappings affect users' performance. We predict that the topographic representation, where adjacent vibration motors map to adjacent locations in space, will facilitate the best performance.

Exploring Harmony Space

We plan to develop a system that uses Holland's Harmony Space system [8, 9] to provide a tactile spatial representation of harmonic structure to musicians learning to improvise. Beginning improvisers typically get stuck on 'noodling' around individual chords from moment to moment and are unable to interact meaningfully with the strategic, longer term harmonic elements, for example, chord progressions and modulations, which are typically essential to higher-level structure in western tonal music, including jazz and much popular music.

Harmony Space draws on cognitive theories of harmonic perception, providing consistent uniform spatial metaphors for virtually all harmonic phenomena, which can be translated into spatial phenomena such as trajectories, whose length, direction and target all encode important information. Thus, Harmony Space enables numerous harmonic relationships to be re-represented in a way that may be more cognitively tractable.

We will use the Harmony Space representation to provide musicians with a tactile representation of the harmonic relationships of music they are currently playing. This will be achieved by having the musicians wear a vest with a 6x8 array of tactile actuators where each actuator will represent a note that the musician is playing. The notes will be identified directly in the case of electronic instruments, or sensed using microphones and pitch trackers in the case of acoustic (monophonic) instruments. We predict that representing pitch movement in this way will facilitate the development of a spatial understanding of musical relationships, which will transfer to improved performance in a wide variety of musical tasks, including improvisation. We will investigate whether performance is linked to the interface becoming transparent.

CONCLUSION

The E-Sense project is taking an interdisciplinary approach to investigating the extended mind, in particular the nature of sensory augmentation. We will use a rapid prototyping approach to build 3 novel tactile interfaces that mediate different sensory modalities (ultrasound, pitch and 'virtual' location). As well as testing the practical utility of these systems, we hope to gain more insight into the conditions under which technologies become transparent as well as gather more evidence for the theoretical viability of the sensorimotor contingency model.

ACKNOWLEDGEMENTS

This research is supported by the Arts and Humanities Research Council grant number: AH/F011881/1.

REFERENCES

1. Bach-y-Rita, P. *Brain Mechanisms in Sensory Substitution*. Academic Press, NY, 1972.
2. Arduino electronics prototyping platform http://www.arduino.cc/ Retrieved June 2008
3. Bird, J., d'Inverno, M. and Prophet, J. Net Work: An Interactive Artwork Designed Using an Interdisciplinary Performative Approach. *Digital Creativity*, 18 (1), (2007), 11–23.
4. Bird, J., Stokes, D., Husbands, P., Brown, P. and Bigge, B. Towards Autonomous Artworks. *Leonardo Electronic Almanac*, (forthcoming).
5. Clark, A. *Natural-Born Cyborgs: Minds, Technologies and the Future of Human Intelligence*. Oxford University Press, NY, 2003.
6. Heidegger, M. *Being and Time*. Harper and Row, NY, 1962.
7. Ho, C., Tan, H. Z. and Spence, C. Using spatial vibrotactile cues to direct visual attention in driving scenes. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8, (2005), 397–412.
8. Holland, S. *Artificial Intelligence, Education and Music*. PhD thesis, IET, The Open University, Milton Keynes. Published as CITE report No. 88, 1989.
9. Holland, S. Learning about harmony with Harmony Space: An overview. In M. Smith and G. Wiggins, (Eds.) *Music Education: An Artificial Intelligence Approach*. Springer Verlag, London, 1994.
10. Kohler, I. The formation and transformation of the perceptual world. *Psychological Issues*, 3, (1964), 1-173.
11. LilyPad sewable electronic components http://www.cs.colorado.edu/ buechley/LilyPad/index.html Retrieved June 2008
12. Maravita, A. and Iriki, A. Tools for the Body (Schema) *Trends in Cognitive Sciences*, 8, (2004), 79–86.
13. O'Regan, J. K. and Noë, A. A Sensorimotor Account of Vision and Visual Consciousness *Behavioral and Brain Sciences*, 24(5), (2001), 939-73.
14. Processing programming language and environment http://www.processing.org/ Retrieved June 2008
15. Ramachandran, V. S. and Blakeslee, S. *Phantoms in the Brain: Probing the Mysteries of the Human Mind*. Fourth Estate, London, 1998.
16. Robinson, D. and Campbell, R. *Contributory Factors to Road Accidents* Transport Statistics: Road Safety, Department for Transport, 2005. http://www.dft.gov.uk/ Retrieved June 2008
17. Rogers, Y. and Muller, H. A Framework for Designing Sensor-Based Interactions to Promote Exploration and Reflection. *International Journal of Human-Computer Studies*, 64 (1) (2005), 1–15.
18. Rogers, Y., Scaife, M. and Rizzo, A. Interdisciplinarity: an Emergent or Engineered Process? In S. Derry, C. D. Schunn and M. A. Gernsbacher (Eds.) *Interdisciplinary Collaboration: An Emerging Cognitive Science*. LEA, (2005), 265–286.
19. Stratton, G. M. Some preliminary experiments of vision without inversion of the retinal image. *Psychological Review*, 3, (1896), 611-617.

SocialSense: A System For Social Environment Awareness

Robert S. Brewer, Samuel R. H. Joseph, Guanghong Yang, Neil Scott, Daniel Suthers
University of Hawaii
Ubiquitous Wireless Applications Group
Laboratory for Interactive Learning Technologies
Department of Information and Computer Sciences
1680 East West Road, POST 309
Honolulu, HI 96822, USA
rbrewer@lava.net

ABSTRACT

SocialSense is a system designed to provide additional social information about nearby people. SocialSense detects Bluetooth devices and uses them to infer the presence of particular people, pulling their profiles and status from online social networking sites. SocialSense differs from existing mobile social awareness systems by integrating live feeds from multiple sources. Information is shown to the user via a head-mounted display, and the user controls the system using buttons mounted on a ring input device or "Magic Ring". The aim is a system that can be used unobtrusively, allowing users to go about face-to-face interactions in a normal manner.

Author Keywords

social networking, wearable computer, presence sharing.

ACM Classification Keywords

H.5.3 [Information Interfaces And Presentation (e.g., HCI)]: Group and Organization Interfaces — Collaborative computing

INTRODUCTION

As people go about their lives, they pass through spaces filled with other people. They will interact with some of these people, but most will be passed by without interaction. One barrier to interaction is unfamiliarity: people are less likely to talk to a stranger they don't know anything about. There is also forgetfulness, such as remembering someone's face but forgetting their name, organizational affiliation, and interests.

This paper describes a system called SocialSense that allows users to be more aware of the social background of people in the environments they inhabit. SocialSense allows the user to explore the profiles and status information of nearby people who have agreed to participate in the system. Profiles are retrieved from an online community site, while status comes from the Twitter microblogging service [14]. Twitter status information consists of a message of up to 140 characters, similar to mobile SMS messages, and provides a potentially dynamic snapshot of a person's current thoughts or activities. The current prototype scans for nearby Bluetooth devices as a proxy for the people in the user's vicinity. The profiles are shown to the user via a head-mounted display (HMD), and the user controls the system using buttons mounted on a ring input device or "Magic Ring". We see the combination of technologies in SocialSense as particularly important. The HMD allows us to display profile icons in the user's peripheral vision to be attended to or ignored based on the user's wishes, as in the eye-q system [3]. The Magic Ring is a deliberately simple input device, designed to allow users to navigate the user interface as easily as possible. While the current SocialSense prototype is quite bulky, we aim to develop a system that can be used unobtrusively, which is important for a system designed to aid social interactions.

For example, a SocialSense user could be walking through a University courtyard filled with people on their way to lunch. As the user is walking, a thumbnail picture of a colleague appears at the edge of their field of view, indicating that the person is nearby. Without this notification, the user might not have noticed the presence of the colleague. Picking them out of the crowd, the user approaches the colleague and asks if they are free for lunch. As they walk to lunch, the user can see their colleague's most recent Twitter status update regarding a paper submission to an upcoming conference. The user is also going to that conference, potentially providing a fertile topic for lunchtime conversation.

RELATED WORK

SocialSense brings together research on location-based social networking systems and alternative input devices.

UbiComp '08 Workshop W1 -- Devices that Alter Perception (DAP 2008)
September 21st, 2008
This position paper is not an official publication of UbiComp '08.

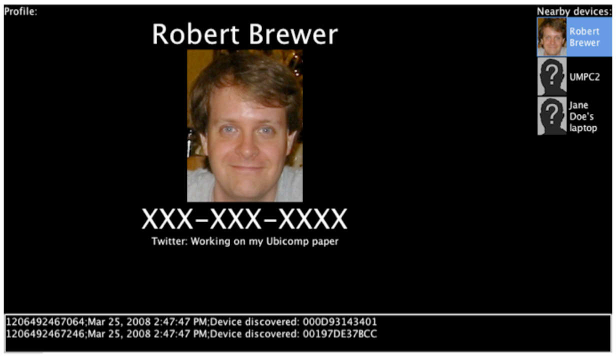
Location-based Social Networking Systems

Social network services such as MySpace and Facebook allow users to create profiles for themselves, such as uploading a picture and specifying friendship links with other users. Commercial systems for mobile and location based social networking services make use of self-reported location (e.g., SocialLight), GPS (e.g., Loopt), and Bluetooth (e.g., MobiLuck) in order to leverage location and context specific social information. All Bluetooth devices are capable of ‘device-discovery’, which allows them to collect information on other Bluetooth devices within 5-10 meters [5]. This information includes a unique Bluetooth MAC address (BTID), device name, and type. The BlueAware system [5] runs in the background on MIDP2-enabled phones allowing them to record and timestamp BTIDs in a proximity log and making them available to other applications. Researchers have been using the BTID patterns to analyze and predict relationships between users and organizational rhythms [5, 13]. Bluscreen is a public advertising system [16] that detects users via their Bluetooth devices and has advertising agents bidding for screen time. Commercial social networking systems such as MobiLuck allow mobile phones to detect all nearby Bluetooth devices, ringing or vibrating when found, supporting message and photo exchange. WirelessRope also uses Bluetooth and supports contact between groups of colleagues at a conference [11]. The Jabberwocky system [12] investigates the “familiar stranger” concept of people who have seen each other in public places on multiple occasions but have never met. The Jabberwocky devices log Bluetooth IDs and no central server is involved, unlike SocialSense.

These systems give us a feel for the possibilities of consumer devices in the mobile social networking field. In addition, there have been many custom social networking applications developed in the wearable computing field including the infamous lovegety [8], GroupWear [2], Smart-Its Friends [7], nTag, and SpotMe. Particularly interesting is the development of systems that incorporate gestural language. For example, iBand [9] is a social networking device that creates connections between two users when they shake hands.

Input Devices

Effective interaction technology is also important when using a head-mounted display and there have been a number of gesture-based interfaces developed including Ubi-finger [17], GestureWrist [15], FingeRing [6], and Twiddler (http://www.handykey.com/). There have been several input devices developed in a ring form factor. FingerSleeve [18] has a six-degree-of-freedom tracker, with which you have ability to sense all movement, and translation and orientation changes. However, it is unsuited for our application because of its size and wire connection, and because SocialSense does not require that level of tracking functionality.



Robert Brewer

XXX-XXX-XXXX

Twitter: Working on my Ubicomp paper

```
1206492467064;Mar 25, 2008 2:47:47 PM;Device discovered: 000D93143401
1206492467246;Mar 25, 2008 2:47:47 PM;Device discovered: 00197DE37BCC
```

Figure 1: SocialSense user interface, showing an abbreviated profile (redacted for privacy)

PROTOTYPE DEVELOPMENT

SocialSense consists of a computer with a Bluetooth adapter that continuously scans for nearby Bluetooth devices. For each Bluetooth device discovered, it contacts a server to see if there is a profile associated with the BTID of the discovered device. If a profile is found, the information is downloaded and added to a list of nearby devices. In addition, if there is a Twitter account associated with the profile, the latest status message is retrieved. Devices that are not associated with a profile are also displayed, but the only information that can be displayed is the name that the device provides (which can sometimes be helpful, such as “Adam Smith’s iPhone”) [10].

Software Implementation

The SocialSense client is written in Java. This decision was made early on because Java allows for cross platform development and deployment. Of particular note is the availability of a cross-platform specification for using Bluetooth with Java, known as JSR 82 (http://jcp.org/en/jsr/detail?id=82).

The user interface is simple by design. It displays the detected users by name and thumbnail image on the right hand side, and the currently selected profile in the center. Log messages are displayed at the bottom of the window showing the status of Bluetooth scans and any errors encountered. Figure 1 shows the user interface.

The interface uses white text on a black background because on some optical see-through HMDs black is transparent thus avoiding unnecessary occlusion of the real world.

To select a device, the user shifts the selection up and down in the list. Moving the selection off the top or bottom of the list causes the profile area to be cleared, allowing the user to focus on his or her physical environment instead of the interface. When a person is selected, that person’s abbreviated profile is displayed, showing their name, picture, phone number and Twitter status. The user can then toggle between an extended profile that displays the person’s full bio and the abbreviated profile.

Currently the server side of SocialSense is implemented in Ruby on Rails as part of the larger disCourse online collaboration system. The ability to associate BTIDs with an individual was added to the existing disCourse profile system. The SocialSense client makes a HTTP request (via WiFi) containing each BTID discovered to the server. If there is a profile associated with a BTID, the server replies with a XML document containing the profile contents, which is then parsed by the client. If the profile has an associated Twitter account, the latest 'tweet' is retrieved from Twitter.

Hardware

The SocialSense prototype runs on a Samsung Q1 UMPC (Ultra Mobile PC). UMPCs are like miniaturized laptops, but they run full versions of Windows. The Samsung model has built-in Bluetooth, WiFi, 2 USB ports, and a VGA port for connecting to the HMD.

We initially used the LitEye LE-750, which is an optical see-through device, for the HMD, but found it too bulky and unsuited for social computing applications. We settled on the Creative Display Systems i-Port as a less obtrusive display. The i-Port consists of a modified pair of Oakley sunglasses with the display mounted onto the right hand side. Unlike the LitEye display, the i-Port is not an optical see-through HMD so it does partially occlude the right eye, but it does not occupy the user's full field of view so it allows some situational awareness.

For input to SocialSense, we developed a "Magic Ring" device to match the simplicity of the user interface. The Magic Ring consists of three small buttons attached to a metal ring, which is attached by wires to a wrist-mounted controller and battery. The wrist-mounted device communicates wirelessly to the receiver module, which attaches to the UMPC via a USB cable. The receiver module appears as a keyboard to the UMPC, and the three buttons send the keystrokes for up arrow, Enter, and down arrow respectively. We are working on an evaluation of the Magic Ring compared to other input devices for common navigation tasks. Figure 2 shows a picture of the device.

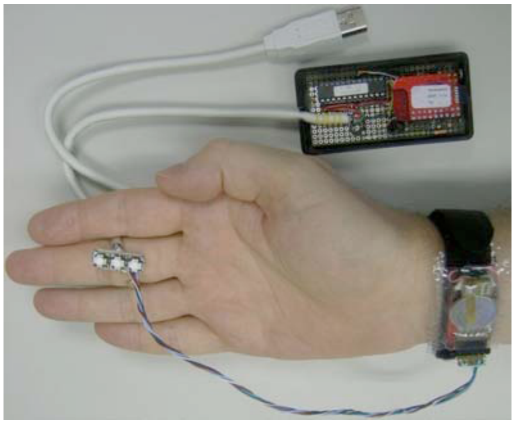


Figure 2: Magic Ring input device

Figure 3: SocialSense hardware being worn

FUTURE WORK

The SocialSense system is still in an early prototype phase, and although we have a working prototype, there are many ways in which it could be improved.

Unobtrusiveness

Significant work is still required before the system can be considered unobtrusive. The head-mounted display is probably the most difficult hurdle towards unobtrusiveness. Current displays are simply too bulky and obvious. While there are many companies working on technologies that they claim will be tiny and practically invisible, only time will tell if these displays live up to the manufacturers' claims.

It may be some time before we can develop a system that can be truly unobtrusive. An alternative approach would be to develop a version of SocialSense for a mobile device like the Apple iPhone. Such a device would be relatively unobtrusive, but it would require a way to make the user aware of nearby people. Given the near ubiquity of Bluetooth headsets, one option would be to have the mobile device "whisper" in the user's ear when someone entered their social space, at which point the user could browse profiles on their mobile device if they wished to.

Beyond Profiles

While profiles from social networking sites can be useful snapshots of a person's identity and interests, they can grow stale if the user does not update them. Updating one's profile does not provide any direct benefit to the user updating the profile; it only helps others. However, there are other sources of data that we can display such blog posts, or FaceBook updates. These information sources, like Twitter, could provide a more up to date indication of what is relevant to the person in question.

The system could even display email messages from the detected individuals that had been sent to the SocialSense user. Such a feature could be very helpful in making sure conversations with colleagues didn't require repetitive explanation of unread emails.

Privacy

With any social networking application, privacy issues are crucial and this is especially true in a mobile wireless environment. The SmokeScreen system [4] allows users to engage in presence-sharing using Bluetooth IDs or WiFi MAC addresses, but provides privacy management using cryptography. SmokeScreen provides a method for presence sharing between strangers using a centralized broker service. Privacy controls can also be on the server side where the user profiles are stored; allowing users to display only limited profile information to users not on their 'buddy list'. The server could also record who retrieved a profile, providing awareness to those being looked up. Critical for privacy is making sure that SocialSense is "opt-in", i.e. you decide if you want to share your profile and who you want to share it with.

Augmented Reality

Azuma and colleagues [1] define an augmented reality (AR) system as one that combines real and computer-generated information in a real environment, interactively and in real time, and registers virtual objects with physical ones. A future AR-enabled version of SocialSense could make the retrieved profiles appear to float above peoples' heads from the perspective of the user wearing the HMD. This would make it obvious who the profiles referred to, but such a feature would require significant advances in AR technology to be practical.

CONCLUSION

We have presented SocialSense, our application for providing context to social situations by sensing Bluetooth devices and displaying nearby user profile and status information. We have developed a prototype using a HMD and the custom Magic Ring input device. The prototype works, but is too cumbersome for routine use. We believe that in time it may be possible to develop an unobtrusive version that displays helpful information about nearby people and we have mapped out several areas for future research.

REFERENCES

1. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., and MacIntyre, B. Recent advances in augmented reality. *IEEE Computer Graphics and Applications* 21, 6 (2001), 34–47.
2. Borovoy, R., Martin, F., Resnick, M., and Silverman, B. (1998) GroupWear: nametags that tell about relationships. In *CHI 98*, ACM (1998), 329-330.
3. Costanza E., Inverso S. A., Pavlov E., Allen R., Maes P., eye-q: Eyeglass Peripheral Display for Subtle Intimate Notifications. In *Proc. of MobileHCI 2006*, (2006), 211–218.
4. Cox, L. P., Dalton, A., and Marupadi, V. SmokeScreen: flexible privacy controls for presence-sharing. In Proc. MobiSys '07. ACM (2007), 233-245.
5. Eagle N. & Pentland. A. S. Reality mining: sensing complex social systems. *Personal Ubiquitous Computing* 10, 4, (2006), 255–268.
6. Fukumoto, M. and Tonomura, Y. Body coupled FingeRing: Wireless wearable keyboard, In *Proc. CHI 97*, ACM (1997), 147-154.
7. Holmquist L.E., Mattern F., Schiele B., Alahuhta P., Beigl M. & Gellersen H.-W. Smart-Its Friends: A Technique for Users to Easily Establish Connections between Smart Artefacts. Proc. *Ubicomp*, (2001), 116-122.
8. Iwatani, Y. Love: Japanese Style. Wired News, 11 Jun 1998.
9. Kanis M., Winters N., Agamanolis S., Gavin A., and Cullinan C. Toward Wearable Social Networking with iBand, In *CHI 2005*, ACM Press (2005), 2–7.
10. Kindberg, T., Jones, T. "Merolyn the Phone": A Study of Bluetooth Naming Practices. Ubicomp 2007 In Lecture Notes in Computer Science 4717 Springer Berlin (2007), 318-335.
11. Nicolai T., Yoneki E., Behrens N. & Kenn H. Exploring Social Context with the Wireless Rope. In *Int'l Workshop on MObile and NEtworking Technologies for social applications*, 2006.
12. Paulos, E. and Goodman, E. 2004. The familiar stranger: anxiety, comfort, and play in public places. In *Proc. CHI 2004*. ACM (2004), 223-230.
13. Perkio J., Tuulos V., Hermersdorf M., Nyholm H., Salminen J. and Tirri H. Utilizing Rich Bluetooth Environments for Identity Prediction and Exploring Social Networks as Techniques for Ubiquitous Computing. In Proc. *IEEE/WIC/ACM Int'l Conf. on Web Intelligence*, IEEE (2006), 137-144.
14. Pontin, J. From many tweets, one loud voice on the Internet. *The New York Times*, April 22, 2007.
15. Rekimoto, J. GestureWrist and GesturePad: Unobtrusive Wearable Interaction Devices, In *Proc. Int'l Symposium on Wearable Computers*, IEEE (2001), 21-27.
16. Rogers, A., David, E., Payne, T. R., and Jennings, N. R. 2007. An advanced bidding agent for advertisement selection on public displays. In *Proc Autonomous Agents and Multiagent Systems*, ACM (2007), 1-8.
17. Tsukada, K. and Yasumura, M. (2002) Ubi-Finger: Gesture Input Device for Mobile Use. In *Proc. APCHI 2002*, 388-400.
18. Zeleznik, R. C., LaViola, J. J. Jr., Feliz, D. A., and Keefe, D. F. Pop Through Button Devices for VE Navigation and Interaction. In *Proc IEEE Virtual Reality*, IEEE (2002), 127-134.

Boxed Ego

Alvaro CASSINELLI and Masatoshi ISHIKAWA
University of Tokyo
alvaro@k2.t.u-tokyo.ac.jp

ABSTRACT

Boxed Ego is a double trap for the Self. A peep-show box waiting in a corner of the exhibition space first captures the curiosity of the observer - and then the observer himself. Although of an artistic flavor, from the research perspective this work is a preliminary experiment on the cognitive (and possible practical) aspects of artificial autoscopy (AS). In order to understand how artificial autoscopy can generate an out-of-body experience (OBE), we embrace the enactive approach to perception [1] and we further hypothesize that the *sense of self*, may be itself a second-order perceptual experience: that resulting not from the exploration of the world based on skillful mastery of the visual, tactile, proprioceptive or auditive sensorimotor contingencies (SMCs), but on exploring/acting on the world with skillful mastery of these SMC *as well as the rules governing the relations (extended in time) between these SMCs*. A first corollary of this hypothesis is that there may be different senses of self: at one extreme, those inextricably linked to each primal sense (and thus experientially ineffable), and at the other extreme, a more abstract sense of self that result from the knowledge of cross-modal contingencies. In between, there may be experiences rendering a more or less unified sense of self, which is precisely why this model seems to us ideal to explain OBEs. A second corollary of this view, is that attentional blindness may also pertain to the sense of self, a testable hypothesis.

Author Keywords

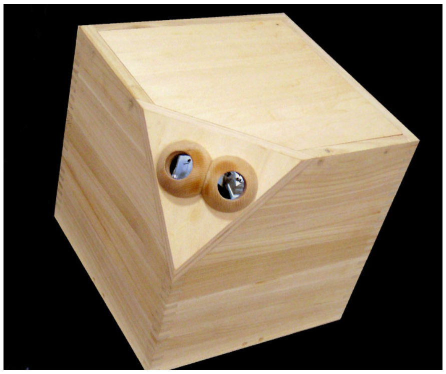
telexistence, out-of-body, autoscopy, self-awareness

ACM Classification Keywords

H.5.1 Multimedia Information Systems — Artificial, augmented, and virtual realities,
H.5.2 User Interfaces — User/Machine Systems

INTRODUCTION

That language and consciousness are inextricably interrelated is not a coincidence since language is a more or less natural formalization of conceptual reasoning, playing a crucial role in the process of self-representation and subjective consciousness [2]. But language alone is not sufficient and surely not even indispensable in order to provide organism self-awareness. How can someone/something incapable of describing knowledge of his/its internal states (even to oneself/itself) be capable of self-awareness? The paradox disappears if one consider that 'description' (internal or external) does not need to be propositional, but can be *enactive* [3].



With this remark in mind, we will leave aside the problem of language-based self-reference, and concentrate instead on enactive forms of self-awareness (as a passing remark, let's note that the ineffable character of enactive knowledge may be responsible for the ineffable part of the sense of self). For one, vision plays a fundamental role in the generation of an egocentric perspective on the world; visual artists have been experimenting in this arena well before science created the right tools or even the proper language capable to describe such phenomena. Self-referential pictures have been around from ten of thousands of years, and artificial mirrors are thousands of years old; however, it's the invention of magnetic recording and closed loop video that opened really new exploratory possibilities. 'Present Continuous Past(s)' by Dan Graham (1974) is perhaps one of the first interactive video-art installations challenging the special vantage point of the audience, and transforming the spectator into its own object of observation. Time delay is purposely used to trick the spectator into the belief that he is seeing a pre-recorded scene unrelated to himself, but then he would slowly gain understanding of his central role in the piece. This calculated spatio-temporal disembodiment brings confusion: as with the Necker cube, the perceptual content is of flipping nature: that of the filmed person being someone else or being oneself. Only very recently these experiments were reproduced in a controlled environment [4]. In this workshop, I would like to foster an informal discussion about the scientific, practical (and of course artistic) potential of this kind of experimentation by describing a media-art installation called 'Boxed Ego' [5].

UbiComp '08 Workshop W1 – Devices that Alter Perception (DAP 2008), September 21st, 2008.

BOXED EGO INSTALLATION

A pair of cameras are aimed towards a small platform on a corner of the exhibition space over which sits a cubic peep-show box. The holes of the peep-box are in fact the eye-pieces of a live-stereoscope. The separation of the video cameras in real space is set to about ten times the real inter-ocular distance, so that the viewer will see a ten times scaled-down version of himself inside an equally miniaturized exhibition space (hyperstereo effect). The box appears empty; however, if the observer talks or breathes, the box readily detects this human prey and traps it in its interior, effectively transforming the observer into its own object of observation. Indeed, a dwarfed, truly three-dimensional version of the observer (peering inside an even smaller box) will slowly materialize (figure 1). Perhaps the main difference between Boxed Ego and other works featuring artificial autoscopy (either in the Media Arts or in the field of experimental psychology [4]) is that (1) the object/subject is perceived truly in 3d, although miniaturized (thus combining autoscopy with micropsia, which are both phenomena that correlate somehow in the medical literature ; (2) the spectator is filmed from behind, and without a time delay it becomes impossible for him to see his own face (this makes the experience very different from that of a mirror or a camera on top of a screen, reminding us of Magritte's famous painting 'La reproduction Interdite'); (3) there is a limited form of correlated tactile feedback (the spectator can grasp the box and see himself grasping it, while at the same time feel the real box his hands); (3) lastly, although not sufficiently compelling in this experiment, the suggested infinite recurrence of observer-observers could potentially generate a sense of multiple body relocation (see below).

The idea behind this installation was to explore, in an artistic way, the links between *curiosity and voyeurism*. While peering inside the box, one can see oneself in every detail, and to a certain extent play with one's own avatar (in particular thanks to some time delay in the video loop). At the same time, one cannot see the other people in the exhibition space (see video in [5]). The installation was exhibited for a week at SonarMatica Media Art festival in Barcelona (2008) with much success. A commentator later reasoned that this could be because 'the theme of self-voyeurism is unsurprisingly very popular with the festival goers.' We agree with this remark (after all, even a simple mirror always retains some magic), but the question remains open: why are we so attracted by these devices? Of course there is a practical aspect to the experience (e.g. tighting your necktie); however, we hypothesize that there is more to this: this sort of setup brings us close to an out-of-body experience which is interesting per se: it give our minds the opportunity to better itself in the mastery of the sensorimotor contingencies in an unusual territory.

THE OUT-OF-BODY EXPERIENCE

Out-of-body experiences (OBEs) are a culturally invariant neuropsychological phenomena that can take a variety of different forms, ranging from seeing one's own body from an elevated visuospatial perspective (the placement of the stereo cameras in the Boxed Ego installation tries to capture this) to the less known 'heautoscopic' hallucination, consisting on perceiving a duplicate of ones body in extrapersonal space [6]. Although the etiology of the OBEs varies widely (organic dysfunctions such as epilepsy, sleep-paralysis, psychological disorders or traumatic experiences but also episodes without a know trigger), direct electrical stimulation of the cortex in pre-operatory brain surgery for intractable epilepsy as well as less invasive experiments (trans-cranial electrical or magnetic stimulation [7]) and fMRI performed during paroxystical hallucinations, all point to the involvement of a very specific area in the brain, namely the *temporo-parietal area* [6].

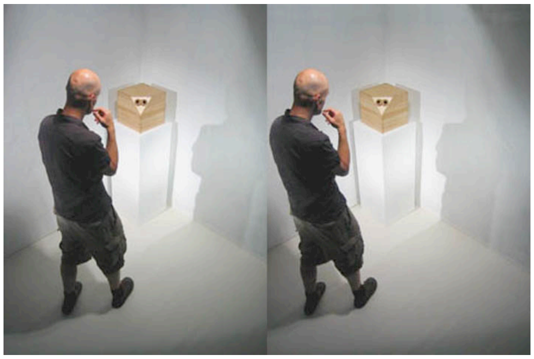


Figure 1. Stereo pair as displayed inside the box (without optics)

Complete distal attribution and OBE

It is interesting to note that although classical OBE imply whole visuospatial relocation in space, it is also possible to have relocated *parts* of the body. This partial relocation is a relatively common occurrence described in the medical literature [9], but also easily reproducible on healthy subjects [10]. It may be argued that 'relocation' of sensation is a normal way of functioning of the sensory-motor apparatus: for any practical purpose, it *must* feel like the sensation is precisely located at the site of stimulation (e.g. on the tip of our finger), instead of, say, inside the head. We always feel *located sensations*, and in particular located in a part of the world that we perceive as 'ours'. *Distal attribution* is the technical term for a very common phenomena, that of situating the stimulus where the action responsible for it is taking place - even if this part is extracorporeal. That's why we feel the texture of paper at the end of the pen, not on your fingers where the force is actually sensed. Distal attribution is exploited in robotic telexistence systems (the user can operate the robot on the same room, or be in another continent for that matter). However, there seems to be a threshold of sensory immersion and sensory-motor correlation that when reached, transforms the fairly common experience described as distal attribution into something qualitatively different: *it elicits a sense of presence in extracorporeal space*. It is therefore tempting to see OBEs as the consequence of a full body relocation in which the experiencer can still see his original body (an experience with an entirely different phenomenology).

SENSE OF PRESENCE: A SENSORIMOTOR ACCOUNT

An ineffable sense of self

As noted in [1], a subset of the 'apparatus-based' sensorimotor contingencies (SMCs) relevant to the sense of vision may derive from sensorimotor laws relative to an 'observer oriented coordinate system'. Learning these laws would provide the system with a rudimentary (enactive) notion of self. For instance, objects (or other people) generate stimuli that can be removed and put back into the visual scene, while sensation about one's own body is always potentially available. Furthermore, some parts of the perceived environment can be controlled at will (i.e. in a manner independent to the motion of the sensory apparatus) while others not (e.g. we don't need to look away in order to hide our own hand). We can generalize this claim as follows: things that are *not ourselves* generate stimuli that can be removed or put back into the *visual, auditive or tactile scenes*, while our own body generates stimuli (including this time proprioceptive information) that cannot be so easily removed. Therefore a sense of self-location is brought by active exploration of the world with (implicit, practical) knowledge of the structure of *egocentric* sensorimotor contingencies. (If the observer was not physically located in a particular place in space, these sensorimotor contingencies would be of a very different nature; perhaps one day a robot with pervasive sensors and actuators - like HALL9000 supercomputer from '2001: Space Odyssey'- will tell us what it's like to have an ubiquitous sense of self).

Sensorial awareness and sense of self

SMCs determined by the character of the 'sensory apparatus' would roughly correspond to the crude character of 'sensation', while those related to the character of the explored objects would form the basis of 'perceived content' [1]. In other words, *awareness* of the character of the experience (is it visual, auditive or something else?) as well as understanding of its content (for the purpose of thought, planning and speech behavior) may be worked out by a concurrent neural mechanism responsible of recognizing and analyzing each particular pattern of SMC. In fact, there may be different levels of 'understanding' (each more or less accessible to consciousness). At the top of the hierarchy, we may have abstract knowledge relative to the occurrence of *some form* of sensory experience, as long as the SMC has some recognizable, familiar structure (perhaps learned late in life). In other words, we may be aware of *being experiencing something* without paying attention to the actual content of the experience. This could contribute to (or even form) a sense of self: if while actively exploring the world, familiar patterns of SMCs appear, then you may not only experience something, but you may experience being a Self experiencing that; if, on the other hand, you fail to recognize any patterns, then you may not just be sense-blind: you may not even experience *being* someone at all.

IDEAS FOR EXPERIMENTS AND PRACTICAL USES

Altering in a controlled way the SMC pattern for a particular sensorial modality may be more or less easy to achieve (the inverted-glass experiment [11] is a classic example). However, altering in a controlled way all the sensorimotor contingencies as well as their inter-relations (including time correlations) may be more difficult to do. To start with, the altering device should be multi-modal. An immersive virtual reality environment could be an ideal setup, but the technology for haptic and proprioceptive actuators is not nearly as developed as auditory or visual displays. For example, while it is easy to set an inverted vision experiment, it is not so easy to conceive –left alone design– a setup for 'inverted haptics': it would mean for instance than when touching something with my right hand, I would feel the object on my left hand.

Attentional self-ness for human computer interfaces

Another interesting consequence of this view is that it should be possible to apply the same principles behind attentional blindness (i.e. *experiential* blindness while retaining sensation) and induce *attentional self-ness*. It turns out that this may be a normal occurrence in everyday life: we do perform repetitive tasks automatically, sometimes without even registering in memory the fact that we did them. (In a sense, we are all *philosophical zombies* from time to time.) However, it would be interesting to be able to control this, perhaps in order to reduce cognitive load from tasks that can be done by a machine and don't need attention from part of the user.

Medical Applications

The temporo-parietal junction seems to be the common lesion site in patients suffering from disturbances of the egocentric spatial-relationship with extrapersonal space (a conditional called *visuospatial neglect*). This is not surprising if we believe the results reported in [8]: this region is in fact very involved in the real-time integration of proprioceptive, tactile, visual and vestibular sensory input, generating a three-dimensional, dynamic representation of the body in space. Therefore, one can wonder if artificially manipulating these inputs may lead to some degree of control over the way the body is represented in space, for therapeutic or at least for palliative care. An example related to this may be the 'revival' of phantom-limbs for the purpose of treating associated pain [9]. Another interesting possibility may be the treatment of higher cognitive dysfunctions, such as dissociative identity disorders; indeed, it has been found that OBEs correlate in people with these disorders [8]. In short, we hypothesize that the availability of a machine through which one is capable of artificially creating and manipulating autoscopic imagery may render a sense of control over otherwise contradictory or poorly organized sensorimotor feedback.

Super mirrors?

Perfectly reflecting surfaces capable of creating an image indistinguishable from reality is a relatively recent human invention that can be traced back to the first century AD [12]. Yet it was a luxury object; Modern ubiquitous mirrors are a much more recent invention Therefore one should be surprised more than not about how comfortably we seem to get along with these artifacts. It is well known that most animals do not pass the 'mirror test', and fall pray once and again to the illusion of reflexions, so one has the right to wonder if our getting used to these ubiquitous reflexions is not because of an intensive exposure in our daily lives (fun house mirrors do make us uncomfortable!). However, since a mirror breaks the natural egocentric visuospatial perspective, one can suspect that their intrusion in the visual field may still disrupt the normal integration of visuospatial information. In fact, researchers have shown that the temporo-parietal region is activated when one tries to mentally superimpose one's body on a front-facing schematic human figures, while the same region is not activated when one observes back-facing characters [8]. It is like the mere idea of seeing oneself from an outside perspective had a special experiential content – everyday mirrors may not be so innocent after all! Perhaps a device that could give finer control of this disruption would be more efficient or safer. This remark is particularly important if one is to consider the use of mirrors on vehicles. A (wearable?) 'autoscopic super mirror' could display a 3d model of the observed/observer as seen from any arbitrary position in extrapersonal space, and this position could be naturally controlled by the user after learning a properly designed artificial SMC scheme that *would not disrupt the sense of self in a way that is counterproductive or dangerous for the task at hand*. In the future this may be achieved by mounting several cameras and reconstructing the scene from an arbitrary point of view. Uses of this could range from 'enhanced mirrors' for dancers that could see their own body from any location during rehearsal, to their use on cars, as an enhancement or substitute of the front and rear mirrors (this can be achieved by collecting images from street cameras or from cameras mounted on other cars, or more simply by using a unique fish-eye camera could be mounted high on the car). Research on telexistence systems is solving part of the problem [13]; indeed, these 'super mirrors' are *autoscopic telexistence* systems.

CONCLUSION AND FURTHER WORK

The system described in this paper tampers with two of the sensory stimuli that seems directly involved in the construction of body self-awareness, namely visuospatial input as well as a limited form of tactile feedback. This experiment does seem to generate a mild form of OBE (or at least the feeling of being in a 'twilight zone' and that without care one can be induced an OBE - and be absorbed by the box). A more objective study is needed in order to asses the efficacy of the illusion, but this was not the goal at this stage of the experiment. In this paper we have deliberately concentrated on a rudimentary notion of the self, one that could account at least for some form of body self-perception. Borrowing the terminology of the sensorimotor contingency model, we may say that being-in-the-body is a way of acting on objects in the world. OBEs would result from the *alteration of normal sensorimotor dependencies as well as cross-modal dependencies*. (This view suggests that synesthesia and out-of-body experiences may be co-morbid phenomena, a view for which there seems to be some medical evidence [14]). If this alteration is consistent in time (something that could be done with the help of 'device that alters perception' more complex than a movable mirror for instance), then one can expect that a functional sense of self could be regained once one comes to grips with the new set of artificial SMCs. This may indeed happens in everyday circumstances. For instance, we usually don't experience any severe disturbance of the sense of self when looking at a mirror, nor is our self disintegrated when playing a first-person shooter game. There may be fundamental reasons for that immunity (such that too few sensorial modalities are involved in these experiments), but it may also be that we have learned enough about these abnormal situations so as to 'flip' the whole set of sensorimotor contingencies, and tune to the one that makes more sense (a bistable form of adaptation similar to the one observed in the limited-time inverted glasses experiment [11]). In any case, it would be interesting to design a device capable of a deeper alteration (although controlled and consistent) of the whole scheme of sensory motor contingencies. A first concrete step would be to include some form of synchronized visuo-tactile stimulation in our own experiment; however, instead of passive stimulation as in [4], it would be interesting if the participant could be himself at the origin of the stimulation. For example, the box could have an opening for a hand, through which the participant would reach the head of his avatar; at the same time, some actuator would touch the real head. Another idea would be to set the whole installation on a moving platform that would tilt as the user tilts the box in his hands, thus instantiating a form of vestibular feedback.

ACKNOWLEDGMENT

The first author would like to thank Arnaud de Grave, Stephane Perrin and Pablo Gindel for inspiring discussions that set the mood for the experiment, as well as to Carson Reynolds for interesting commentaries.

REFERENCES

1. A. Noe, *Action in Perception*, The MIT Press (2004).
2. D. Dennett, *Consciousness Explained*, Penguin (1991).
3. F. J. Varela et al., *The Embodied Mind*, MIT (1991)
4. H. H. Ehrsson, *The Experimental Induction of OBEs*, Science, 317(5841): 1048 (2007)
5. www.k2.t.u-tokyo.ac.jp/members/alvaro/boxedEgo
6. O. Blanke and G. Thut, *Inducing OBEs*, Ch.26, Tall Tales about Mind and Brain, (2006)
7. CM. Cook and MA. Persinger, *Experimental induction of the "sensed presence"*. Percept. Mot. Skills. 85(2):683-93 (1997)
8. O. Blanke, *OBEs: Psychological and neurological characteristics*, Proc. 5th Symp. of the Bial Foundation.
9. Ramachandran, V. S. and S. Blakeslee, *Phantoms in the brain*, William Morrow Co. (1998)
10. H. H. Ehrsson et al., *Touching a Rubber Hand*, J. of Neurosc. 25(45):10564-10573 (2005)
11. J. G. Taylor, *Behavioral Basis of Perception*, Yale Univ. Press, (1962)
12. S. M. Bonnet, *The Mirror: A History*, Routledge (2001)
13. K. Watanabe et al., *TORSO: completion of egocentric telegnosis system*, SIGGRAPH (2007)
14. HJ. Irwin, Correspondence. J Soc Psych Res, 51:118-120, (1981)

Fear tuners – Prostheses for instincts

Susanna Hertrich
Royal College of Art
Kensington Gore, SW7 2EU, London, UK
+44.798.308 7004

ABSTRACT

This paper concerns "Fear Tuners", a critical design project that was initiated at the Royal College of Art in 2008. In this paper, I argue that our bodies are equipped with a sensory system that only allows us to detect immediate dangers, for example it helps us to decide where to tread and what to eat. This system though is not suitable to sense the abstract and global dangers that occur in our highly complicated world.

Fear Tuners brings forward the arguments that people are in need of tools to help them sense global and abstract dangers. As a response to the problem, this project explores the potential use of wearable devices as prostheses for those missing instincts. The paper suggests using the skin as an interface to stimulate a physical sensation resulting into a mental state of increased awareness, whenever a deferred danger occurs.

Author Keywords

Augmented cognition, prosthetic design, haptics, wearables, critical design, device art.

ACM Classification Keywords

B.4.2. Input/Output Devices, H.5.m. Information Interfaces and presentation, K.4.1.c Ethics

INTRODUCTION

This paper concerns "Fear Tuners", a critical design project that began life in the Design Interactions Department at the Royal College of Art in 2008. It is a project of *design research*, practiced from the perspective of artist-designers. Fear Tuners stands in the tradition of critical design. This approach aims to open new spaces for designers and means to provide an alternative method to design, in contrast to focusing on the factors of ‘usability’ or commercial viability of a product, service or system. Embodying different values into the designs triggers a debate on the impact of specific technologies that comes with these. The designs can be seen as a manifestation of people’s hopes and fears in relation to those technologies [6].
The Fear Tuners objects are wearable, functional devices, which also stand in the tradition of *device art.* This classification defines artworks that consist of a hardware, which is specifically designed to realize a particular concept. The functional and visual design aspects of these objects make an essential part of the artwork [8].

UbiComp ’08 Workshop W1 – Devices that Alter Perception (DAP 2008)
September 21st, 2008
This position paper is not an official publication of UbiComp ’08.

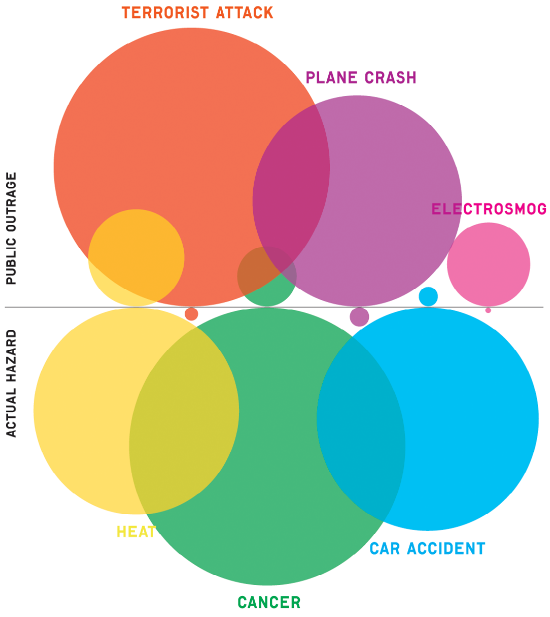


Figure 1. Risk perception and actual hazards

BACKGROUND

The project arose from the insight of being unable to assess the threats, dangers and risks that we are faced with in today's complicated world. Technologies have greatly reduced some of the biggest risks of humankind, yet our modern life seems to spawn a whole new array of abstract threats and fears [1]. Creating a common feeling of "being at risk" has become a popular political method as well it is widely exploited in mainstream journalism [5]. The consent to a common fear in a community can result into a more cohesive society and the choice to be aware of a danger is often meant to conform a specific way of life [3]. Generally, it can be observed that people seem to be unable to differentiate between mere panic mongering and the real threats that surround them. For example, we can register a massive media outrage on minor or non-existing threats (e.g. bird flu, MMR vaccine), and a neglect of many serious risks, such as old age poverty related to non-functional pension schemes (Figure 1).

HUMAN SENSES AND ABSTRACT DANGERS

Our hard-wired sense apparatus is not suitable to sense the modern dangers in an array of fear stories. We are only hardwired to deal with sudden or physical dangers, such as approaching cars, burning fires or rotten food. But we do not have the instincts to sense the abstract and deferred dangers that have a huge effect on our daily lives, like stock market crashes and the rising oil price.

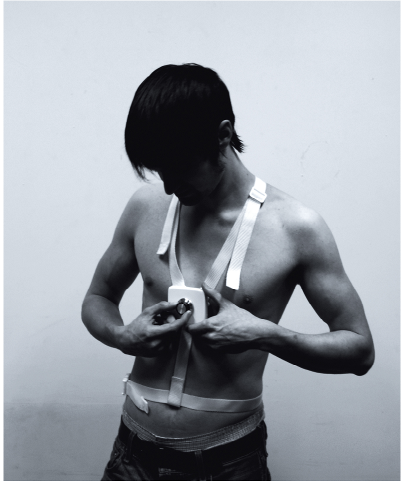


Figure 2. Fear Tuners – Form prototype

I propose to face this inability with the implementation of wearable devices (Figure 2) as prostheses for these instincts to be able to sense the deferred and abstract dangers of today.

SKIN AS INTERFACE

When we sense a physical danger, a set of bodily reactions comes into action. We can feel cold shivers that run down our spines, get goose bumps, sweaty hands, our neck hair raises and we start to tremble. The most extreme of these reflexes is the so-called 'fight or flight response' that jumps into action, whenever we are faced with a sudden attack [2]. In this state, our pupils have narrowed and we have lost peripheral vision, we have an accelerated heart and lung activity, and nutrient has been released to our muscles, among many others, to get us ready for action [7]. None of these physical manifestations is voluntarily chosen or the outcome of an intellectually driven thought process. Instead, they are the immediate reflexes to an instinct sensing danger.

These processes are hard-wired into our bodies as a result of evolution, even though we rarely encounter emergencies that require physical effort.

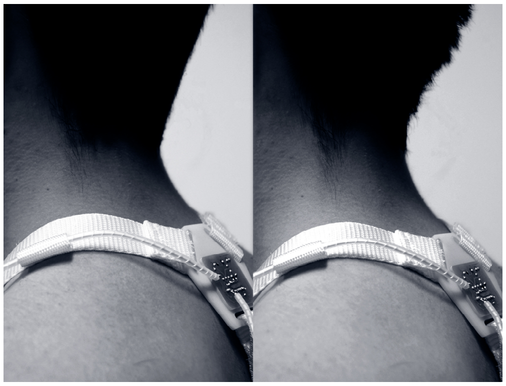


Figure 3. Raised neck hair and increased alertness through physical stimulation

Fear Tuners as prostheses for instincts proposes to use the skin as an interface to stimulate similar physical sensations (Figure 3), as described in the preceding paragraph. Fear Tuners create an equally immediate and intense experience.

Referring on the concept of body-to-emotion-feedback, by Paul Ekman, who describes how voluntary facial actions are capable to generate changes in both autonomic and central nervous system, I propose that wearing Fear Tuners will similarly result into changed mental state. By inducing a set of physical reactions normally related to fear, such as raising a person's neck hair or generating cold shivers and goose bumps, a state of increased awareness will be generated [4].

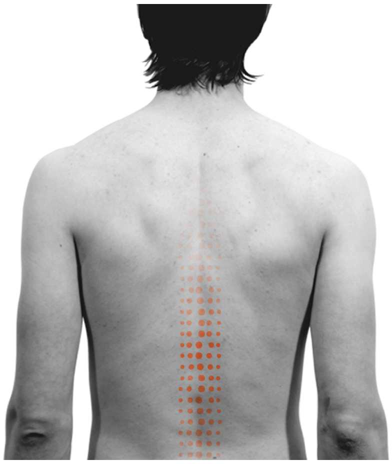


Figure 4. Visualization – Stimulation of cold shivers related to the current inflation rate

AUGMENTATION OF HUMAN INSTINCTS TO PERCEIVE GLOBAL DATA

Fear Tuners are wearable devices, which act directly on the skin. Wireless technology links them to a piece of software that harvests the internet for related data streams, e.g. stock market data, oil price etc. Whenever a severe change in data occurs, the device passes on a sensation to the wearer.

Presenting the information in form of physical stimuli, rather than intellectual (textual and image based information), allows the Fear Tuners wearer to focus the center of his or her attention on other things. The wearer can completely process Fear Tuners' signals in the background of awareness. This form of ambient information presentation engages the senses and thus results into a subtle, yet intense experience that does not disrupt the wearers daily routine [10].

In the process of exploring suitable sensations, I was investigating different actuators, such as solenoids and vibration motors, peltier pumps and electrical deep tissue stimulation aiming to create cold shivers (Figure 4), goose bumps, raised neck hair and hot stings. I also looked into possibilities of exploiting the phenomenon of somatosensory illusions [9].

I identified five key scenarios, Disasters, Financial, Health, Personal and Technology, in which Fear Tuners would act as an 'artificial sixth sense' in the form of a device.

CONCLUSIONS

At present, Fear Tuners exist as a series of technical experiments, form prototypes, a video scenario and booklet. They were presented as part of my thesis at the Royal College of Art graduation show. I am hoping to bring the project to a next level, in which the preceding research and experimentation in form and function would be combined to create to a fully functional prototype. For this next step, I am looking for collaboration partners from a different background other than design.

ACKNOWLEDGMENTS

I thank Fiona Raby, Tony Dunne and James Auger who guided and helped to develop the project at the Royal College of Art and Carson Reynolds for his valuable advice and inspiration.

REFERENCES

1. Aldersey, W., Briscoe, S., *Panicology*, Viking/Penguin, London, UK, pp. XIV-XVIII. 2008.
2. Cannon, W. B., *The wisdom of the body*, New York, NY, Norton, 1932.
3. Douglas, M., Wildavsky A., *Risk and Culture*, University of California Press, Berkeley, CA, USA, 1983.
4. Ekman, P., E. T. Rolls, D. I. Perrett and H. D. Ellis, Facial Expressions of Emotion: An Old Controversy and New Findings [and Discussion], *Philosophical Transactions: Biological Sciences*, Vol. 335, No. 1273, Processing the Facial Image, pp. 63-69, 1992.
5. Gardner, D., *Risk*, Virgin Books, London, UK, 2008.
6. Gaver W., Dunne, T. & Pacenti, E., Cultural Probes, *Interactions*, pp. 24-25, 1999.
7. Jansen, A. S., Nguyen, X. V., Karpitskiy, V., Mettenleiter, T. C., and Loewy, A. D., Central command neurons of the sympathetic nervous system: basis of the fight-or-flight response. *Science* (New York, N.Y.), 270(5236): 644-646, 1995.
8. Kusahara, M., Device Art: A New Approach in Understanding Japanese Contemporary Media Art. In *Mediaarthistories*, ed. Oliver Grau. MIT Press, Boston, MA, USA, p. 288, 2007.
9. Sherrick, C. E. & Rogers, R., Apparent haptic movement, *Perception & Psychophysics*, Vol.1, pp. 175-180, 1966.
10. Wisneski, C., Ishii, H., Dahley, A., Gorbet, M., Brave, S., Ullmer, B., and Yarin, P, Ambient displays: Turning architectural space into an interface between people and digital information. Volume 1370 of *Lecture Notes on Computer Science,* pages 22-32., 1998.

Gesture recognition as ubiquitous input for mobile phones

Gerrit Niezen **Gerhard P. Hancke**
University of Pretoria
Lynnwood Road, Pretoria,
0002, South Africa
{gniezen, g.hancke}@ieee.org

ABSTRACT

A ubiquitous input mechanism utilizing gesture recognition techniques on a mobile phone is presented. Possible applications using readily available hardware are suggested and the effects of a mobile gaming system on perception is discussed.

Author Keywords

ubiquitous computing, accelerometers, gesture recognition, optimization, human-computer interfaces

ACM Classification Keywords

B.4.2 Input/Output Devices, H5.m. Information interfaces and presentation

INTRODUCTION

Mobile phones are the most pervasive wearable computers currently available and have the capabilities to alter and manipulate our perceptions. They contain various sensors, such as accelerometers and microphones, as well as actuators in the form of vibro-tactile feedback. Visual feedback may be provided through mobile screens or video eye wear.

Dynamic input systems in the form of gesture recognition are proving popular with users, with Nintendo's Wii being the most prominent example of this new form of interaction, that allows users to become more engaged in video games [1]. The video game experience is now affected not only by timing and pressing buttons, but also by body movement.

To ensure a fast adoption rate of gesture recognition as an ubiquitous input mechanism, technologies already available in mobile phones should be utilized. Features like accelerometer sensing and vibro-tactile feedback are readily available in high-end mobile phones, and this should filter through to most mobile phones in the future.

Hand gestures are a powerful human-to-human communication modality [2], and the expressiveness of hand gestures also allows for the altering of perceptions in human-computer interaction. Gesture recognition allows users to perceive their bodies as an input mechanism, without having to rely on the limited input capabilities of current mobile devices. Possible applications of gesture recognition as ubiquitous input on a mobile phone include interacting with large public displays or TVs (without requiring a separate workstation) as well as personal gaming with LCD video glasses.

The ability to recognize gestures on a mobile device allows for new ways of remote social interaction between people. A multiplayer mobile game utilizing gestures would enable players to physically interact with one another without being in the same location. Gesture recognition may be used as a mobile exertion interface [3], a type of interface that deliberately requires intensive physical effort. Exertion interfaces improve social interaction, similar to games and sports that facilitate social interaction through physical exercise. This may change the way people perceive mobile gaming, as it now improves social bonding and may improve overall well-being and quality of life.

Visual, auditory and haptic information should be combined in order to alter the user's perceptions. By utilizing video glasses as visual feedback, earphones as auditory feedback and the mobile phone's vibration mechanism as haptic feedback, a pervasive mobile system can be created to provide a ubiquitous personal gaming experience. Gesture recognition is considered as a natural way to interact with such a system.

Gesture recognition algorithms have traditionally only been implemented in cases where ample system resources are available, i.e. on desktop computers with fast processors and large amounts of memory. In the cases where a gesture recognition has been implemented on a resource-constrained device, only the simplest algorithms were considered and implemented to recognize only a small set of gestures; for example in [5], only three different gestures were recognized.

We have developed an accelerometer-based gesture recognition technique that can be implemented on a mobile phone. The gesture recognition algorithm was optimized such that it only requires a small amount of the phone's resources, in order to be used as a user interface to a larger piece of software, or a video game, that will require the majority of the system resources. Various gesture recognition algorithms currently in use were evaluated, after which the most suitable algorithm was optimized in order to implement it on a mobile phone [6]. Gesture recognition techniques studied include

UbiComp '08 Workshop W1 – Devices that Alter Perception (DAP 2008)
September 21st, 2008

hidden Markov models (HMMs), artificial neural networks and dynamic time warping. A dataset for evaluating the gesture recognition algorithms was gathered using the mobile phone's embedded accelerometer. The algorithms were evaluated based on computational efficiency, recognition accuracy and storage efficiency. The optimized algorithm was implemented in a user application on the mobile phone to test the empirical validity of the study.

CURRENT IMPLEMENTATIONS

Choi et al. [7] used accelerometer data acquired from a mobile phone's built-in accelerometer. They were able to recognize digits from 1 to 9 and five symbols written in the air. During their experimental study, they were able to achieve a 97.01% average recognition rate for a set of eleven gestures. The recognition rate was cross-validated from a data set of 3082 gestures from 100 users. This was done using a Bayesian network based approach, with gesture recognition done on a PC connected to the mobile phone.

Pylvänäinen [8] employed an accelerometer-based gesture recognition algorithm using continuous HMMs, with movements recorded using an accelerometer embedded in a mobile phone, but gesture recognition was still performed on a desktop PC. A left-to-right HMM with continuous normal output distributions was used. The performance of the recognizer was tested on a set of 10 gestures, 20 gesture samples from 7 different persons, resulting in a total of 1400 gesture samples. Every model for each of the 10 gestures had 8 states. 99.76% accuracy was obtained with user-independent testing. Pylvänäinen argued that an extensive set of gestures (i.e. more than 10) becomes impractical due to users having to learn all the different gestures.

With gesture recognition one should distinguish between postures, involving static pose and location without any movements; and gestures, involving a sequence of postures connected by continuous motions over a short time span [2]. Crampton et al. [1] developed an accelerometer-based multi-sensor network to recognize both postures and gestures. The wearable sensor network detects a user's body position as input for video game applications, providing for an immersive game experience. Mahalanobis distance is used as a nearest-neighbour means of classification. This improves on using Euclidian distance as a metric, as it takes into account the correlations of the data set and is scale-invariant. They argue that the more accelerometers are used, the more accurately gestures and poses can be differentiated. This should be taken into account when developing a gesture-based system, and is discussed further later in the paper.

Current accelerometer-based motion-sensing techniques in mobile phones are either based on tilt or orientation, allowing for simple directional movement control in games. Camera-based methods for gesture recognition are also becoming more popular. A company called GestureTek [19] enables mobile phones with built-in cameras to be used as motion-sensing devices. In the case of camera-based computer vision algorithms, the necessary image processing can be slow, which creates unacceptable latency for fast-moving video games and other applications [20]. Camera-based sensors are also deemed power-hungry, which is a problem considering that the amount of power consumed during operation is of utmost importance in a mobile device.

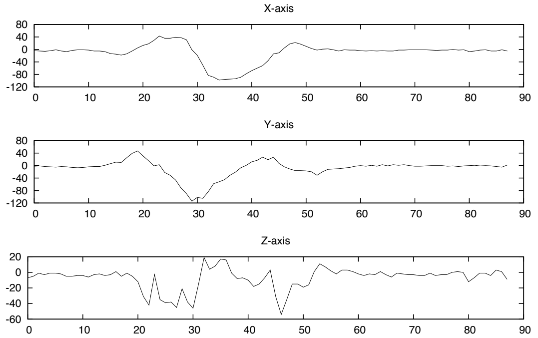


Figure 1. Raw sensor data sampled from the Nokia N95's accelerometer

IMPLEMENTATION AND RESULTS

In [9], we describe how various gesture recognition techniques were evaluated, after which the most suitable algorithm was optimized in order to implement it on a mobile device. We make use of the Dynamic Time Warping (DTW) algorithm, introduced by Sakoe and Chiba [10] in a seminal paper in 1978. The DTW algorithm used was originally implemented in C by Andrew Slater and John Coleman [11] at Oxford University Phonetics Laboratory. The DTW algorithm non-linearly wraps one time sequence to match another given start and end point correspondence.

Sensor data was collected using a Nokia N95's embedded 3-axis STMicroelectronics LIS302DL accelerometer. The Symbian 3rd Edition SDK's Sensor API was used to gather raw sensor data using an interrupt-driven sampling method. The data was filtered using both a digital low-pass filter (LPF) and a high-pass filter (HPF). In figure 1 the raw sensor data gathered from the mobile phone's accelerometer is shown for all the three axes.

A total of 8 gestures with 10 samples per gesture were collected. As the DTW algorithm is essentially a type of template-matching technique, only one training sample per gesture was required for the DTW algorithm to perform the gesture recognition correctly. The 8 gestures used in this study can be observed in figure 2. The gestures used were obtained from a study done by Bailador et al. [12]. The DTW algorithm was able to correctly classify a total of 77 of the 80 samples, for an overall accuracy of 96.25%. The algorithm was optimized [9] for the mobile phone and the recognition time was reduced from around 1000 ms to under 200 ms.

The gesture recognition algorithm was ported to the mobile device by making use of Nokia's Open C platform [13]. Open C is a set of POSIX libraries to enable standard C programming on Symbian Series 60 devices.

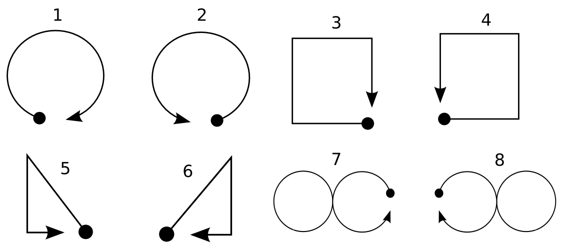


Figure 2. Gestures used in this study

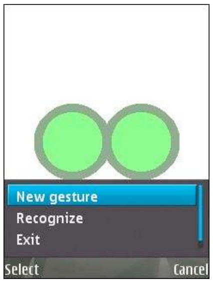


Figure 3. User application running on the mobile device

A user application was implemented to test the real-world functionality of the gesture recognition algorithm. The user application was developed in the Python programming language and executed on the mobile phone using Nokia's Python for Series 60 (S60) version 1.4.1 utilities [14]. Using Python allows one to rapidly prototype a graphical user interface (GUI) and other functionality by making use of the built-in APIs to provide, for example, sound and graphics capabilities. An example of the user application running on the mobile device is shown in figure 3.

The gesture recognition algorithm (written in C) was linked into the Python program as a dynamically linked library (DLL). Wrapper code was created for the C algorithm in order to link it into the Python program. The user application was converted into a standalone Python program on the Symbian device through the Ensymble developer utilities for Symbian S60 [15]. It can also run as a script in the Python for S60 shell.

To have the system learn a new gesture, the user can select the `New Gesture` command from the pop-up menu. When the user starts moving the phone, the application records the gesture until the phone stops moving. The recorded gesture is then stored as a reference gesture on the phone. To recognize a gesture, the user selects the `Recognize` command from the pop-up menu. The application records the test gesture as soon as the user starts moving the phone. When the device stops moving, the application executes the gesture recognition algorithm and displays the recognized gesture as a graphic on the screen.

Nokia's Python for Series 60 does not provide built-in support for vibro-tactile feedback, but third-party utilities have been developed to overcome this. For Series 60 3rd Edition devices (like the Nokia N95) a third-party module called `misty` provides vibration support, and for Series 60 2nd Edition an earlier package called `miso` was developed. These capabilities will probably be added to the Nokia Python library in future.

Haptic feedback was added to the user application by utilizing the vibro-tactile capabilities of the mobile phone when a gesture is recognized. Visual feedback is provided by displaying a graphic of the gesture on-screen. Auditory feedback was added by having the recognized gesture spoken out loud using the text-to-speech functionality of the Nokia Python Audio API.

Personal media viewers, such as the Myvu Crystal [16], allow for a full-screen mobile viewing experience. When combined with a mobile phone such as the Nokia N95 with an embedded accelerometer, our gesture recognition algorithm and a mobile game, the pervasive mobile gaming system as described in the previous sections becomes possible. The Myvu glasses can be connected to the Nokia N95 via the Nokia AV connector, a 3.5 mm stereo headphone plug.

It is envisioned that personal media viewers such as the Myvu will enable mobile gesture-based gaming opportunities until true see-through head mounted displays become less expensive. With the Myvu video glasses it is possible to look above or below the screen, which allows one to walk around. This makes it possible to use the video glasses for urban gaming, or other applications where the user is required to physically walk around while still wearing the video glasses.

Another possible application would be body mnemonics, an interface design concept for portable devices that uses the body space of the user as an interface [17]. Different body positions may be used as markers to remember computational functionality or information such as phone book entries. For example, the user might move the mobile phone to the shoulder or head to access a specific sub-menu or program on the phone. Continuous audio or tactile feedback relating to the user's motion or gesture trajectories may be provided. It is believed that this kind of tightly coupled control loop will support a user's learning processes and convey a greater sense of being in control of the system [18]. User interfaces or functions can now be logically or emotionally mapped to the user's body, completely changing the perception of interacting with a mobile device.

CONCLUSION

Gestures can change the way we interact with computers and mobile devices. This is evident in new user interfaces such as the multi-touch interface introduced by the Apple iPhone. The multi-touch interface adds motion gaming capabilities to the iPhone, albeit in a different sense than using accelerometer-based gesture recognition. This paper describes a cost-effective mobile system that can be implemented with readily available hardware and realizable software on a mobile phone. An optimized gesture recognition algorithm that require minimal resources was described and implemented on a mobile phone.

Accelerometer-based techniques have an advantage above camera-based techniques, i.e. that computationally intensive calculations are not required for accurate movement information, as measurements are directly provided by the sensors. Sensor-based techniques also have the advantage in that they can be used in much less constrained conditions and are not reliant on lighting conditions or camera calibration [21].

To provide a more immersive experience, wireless video glasses may be developed that does away with cumbersome cabling. For the video glasses to be connected to a mobile phone, the wireless technologies used will most probably have to be Bluetooth or Wi-Fi, as these technologies are already available in mobile phones. This is an avenue for further exploration, since as of this writing no true wireless video glasses have been developed.

Possible pitfalls for gesture recognition in mobile phones include user acceptability: Will a user feel comfortable waving his or her arms around in a public space? Haptic feedback is also important for user acceptance. The Nintendo Wii, for example, incorporates this by providing both auditory and vibro-tactile feedback when performing a gesture. A user must know the set of gestures that a system recognizes and gestures requiring high precision over a long period of time can cause fatigue. Therefore the gestures must be designed to be simple, natural and consistent. If the gestures prove to be tiring or strenuous, any possibility of altering the user's perceptions will be limited.

When only one accelerometer is used, the accuracy in detecting the various gestures is reduced. With the Nintendo Wii, for example, the basic motions it detects can easily be cheated with partial movement [1], which reduces the immersive perception of a video game. Utilizing multiple accelerometers increases accuracy at additional cost. Adding additional accelerometer-based sensing devices to a mobile gaming system should not be technically complex, as Bluetooth may be used for communicating with the mobile phone.

Location-based games, also known as urban gaming, can utilize a mobile phone's GPS receiver to provide a realistic, augmented reality-type gaming experience. This may be combined with the methods described in this paper to improve even further on the alteration and modification of the user's perceptions. Hand gestures can also be used in 3D virtual environments to provide a more natural and immersive user experience [2], truly altering users' perceptions in viewing and experiencing their environment.

REFERENCES

1. Crampton, N., Fox K., Johnston H. and Whitehead A. Dance Dance Evolution: Accelerometer Sensor Networks as Input to Video Games. In *Proc. IEEE HAVE 2007*, 107-112.
2. Chen Q., Petriu E.M. and Georganas, N.D. 3D Hand Tracking and Motion Analysis with a Combination Approach of Statistical and Syntactic Analysis. In *Proc. IEEE HAVE 2007*, 56-61.
3. Mueller F., Agamanolis S. and Picard, R. Exertion interfaces: sports over a distance for social bonding and fun. In *CHI '03: Proc. SIGCHI Conf. on human factors in computing systems 2003*, 561-568.
4. Khronos Group. OpenGL ES Overview. http://www.khronos.org/opengles/.
5. Feldman, A., Tapia, E.M., Sadi, S., Maes, P. and Schmandt, C. ReachMedia: On-the-move interaction with everyday objects. In *Proc. IEEE ISWC 2005*, 52-59.
6. Niezen G. The optimization of gesture recognition techniques for resource-constrained devices. M.Eng. thesis, University of Pretoria, South Africa, 2008.
7. Choi, E., Bang, W., Cho, S., Yang, J., Kim, D., and Kim, S. Beatbox music phone: gesture-based interactive mobile phone using a tri-axis accelerometer. In *Proc. IEEE ICIT 2005*, 97-102.
8. Pylvänäinen, T. Accelerometer Based Gesture Recognition Using Continuous HMMs. In *LNCS: Pattern Recognition and Image Analysis*, Springer-Verlag (2005).
9. Niezen G. and Hancke G.P. Evaluating and optimising gesture recognition techniques for mobile devices. Int'l J. Human-Computer Studies. Elsevier (submitted June 2008).
10. Sakoe, H. and Chiba, S. Dynamic programming algorithm optimization for spoken word recognition. In *IEEE Trans. Acoustics, Speech, and Signal Processing*, 26 (1), 43-49.
11. Coleman, J. Introducing speech and language processing. Cambridge University Press, Cambridge, UK, 2005.
12. Bailador, G., Roggen, D., Tröster, G. and Triviño, G. Real time gesture recognition using Continuous Time Recurrent Neural Networks, In *Proc. Int. Conf. Body Area Networks 2007.*
13. Nokia Research Center. Open C: Standard-based Libraries for Symbian-based Smartphones. http://opensource.nokia.com/projects/openc/.
14. Nokia Research Centre. Python for S60. http://opensource.nokia.com/projects/pythonfors60/.
15. Ylänen, J. The Ensymble developer utilities for Symbian OS. http://www.nbl.fi/∼nbl928/ensymble.html.

16. Myvu Corporation. Myvu Crystal. http://www.myvu.com/Crystal.html.

17. Ängeslevä J., Oakley I., Hughes S. and O'Modhrain, S. Body mnemonics - portable device interaction design concept. In *UIST - Adjunct Proc. ACM Symposium on User Interface Software and Technology 2003*.

18. Strachan S., Murray-Smith R., Oakley I. and Ängeslevä J. Dynamic Primitives for Gestural Interaction. In *LNCS: MobileHCI*, Springer-Verlag (2004).

19. GestureTek Mobile. http://www.gesturetekmobile.com.

20. Geer, D. Will gesture recognition technology point the way? *IEEE Computer*, 37(10), 2004, 20-23.

21. Chambers, G.S., Venkatesh, S., West, G.A.W. and Bui, H.H. Hierarchical recognition of intentional human gestures for sports video annotation. In *Proc. Int. Conf. on Pattern Recognition*, 1082-1085.

CREATION OF SYMPATHETIC MEDIA CONTENT

Stéphane Perrin
perrin.japan@gmail.com

Giuseppe Riva
Catholic Univ. of Milan
giuseppe.riva@unicatt.it

Alvaro Cassinelli
University of Tokyo
cassinelli.alvaro@gmail.com

ABSTRACT

Taking ground in the *enactive view*, a recent trend in cognitive science, we propose a framework for the creation of sympathetic media content. The notion of sympathetic media content is based on two concepts: synesthetic media and empathic media transmission.

Synesthetic media is media that make use of multiple and alternative senses. The approach is to reconsider traditional media content from a different perceptual point of view with the goal of creating more immersive and affective media content. Empathic media transmission will consist in encoding the emotional content of media into multi-sensory signals. The encoded emotions are then mediated to the audience through actuators that provide the physical manifestation of the multi-sensory information.

The two points, synesthetic media and empathic transmission, are addressed through the study of the relation between senses and emotions and the development of suitable methods for encoding emotions into multiple senses, in the frame of an efficacious transmission of emotions to the audience. The extraction of emotional information from media and the conception of a wearable, unobtrusive device are considered too. It is claimed that such a framework will help the creation of a new type of media content, ease the access to more immersive and affective media, and find applications in numerous fields.

Author Keywords

media, enaction, emotion, sensors, perception, senses

ACM Classification Keywords

H.5.1 Multimedia Information Systems — Artificial, augmented, and virtual realities,
H.5.2 User Interfaces — User/Machine Systems

INTRODUCTION

An emerging trend in cognitive science is the *enactive view* [6, 7] of sensorimotor knowledge. In this approach, perceiving is to understand how sensory stimulation varies as we act. In particular it implies the *common coding theory* [8]: actions are coded in terms of the perceivable effects they should generate. More in detail, when an effect is intended, the movement that produces this effect as perceptual input is automatically activated, because actions and their effects are stored in a common representational domain.

The underlying process is the following [9, 10]: first, common event representations become activated by the perceptual input; then, there is an automatic activation of the motor codes attached to these event representations; finally, the activation of the motor codes results in a prediction of the action results in terms of expected perceptual events. The enactive view and its corollaries support the concepts of:

- Synesthetic media,
- Empathic media transmission.

After a detailed definition of the new notion of sympathetic media as synesthetic media combined to an empathic transmission, its practical implementation is discussed. Namely, the way to encode emotions and to transmit them through multi-sensory channels is presented as well as the design of a device to achieve this aim.

DEFINITION OF SYMPATHETIC MEDIA

Sympathetic media is the combination of synesthetic media and empathic transmission.

Synesthetic Media:

In cognitive science, synesthesia (Greek, syn = together + aisthesis = perception) is the involuntary physical experience of a cross-modal association. That is, the stimulation of one sensory modality reliably causes a perception in one or more different senses. Specifically it denotes the rare capacity to hear colors, taste shapes, or experience other equally startling sensory blendings whose quality seems difficult for most of us to imagine. A synesthete might describe the color, shape, and flavor of someone's voice, or seeing the color red, a synesthete might detect the "scent" of red as well. Transmission of emotions (for an enactive view on emotions, see [11]; on vision, see [12]), tones, moods or feelings intrinsically contained in media or that a creator intends to transmit via a media to an audience relies heavily on only two senses, the audition (music or speech) and the vision (images or text). On the contrary, human communication relies on a wide range of senses. Moreover, this reliance on only two senses fails in some cases to convey sufficient information to break cultural barriers or to reach audiences with sensory disabilities. The efficiency of information transmission, including emotions [13], can be limited due to an overloading of the visual and aural channels, for example by textual information such as subtitles that is perceived through vision and imply a cognitive effort. The idea of using alternative sensory channels to create more immersive, affective or realistic context and content for the audience is not new, especially in the fields of Ambient Intelligence [14], Immersive, Perceptual or Affective Computing [15], and Human-Computer Interaction [16]. To take an example, most Virtual Reality (VR) rooms include several kinds of sensory outputs (wind [17], scent, force, haptic or temperature [18]) other than vision or audition. Nonetheless, most of these works remain not easily accessible to audiences. Either because of their bulky nature (dedicated spaces for VR) or because there is no seamless integration of the extra sensory information in the media that contains them. Moreover, these works are mostly dedicated, and somehow limited, to re-create perceptual sensations identical to the ones that are virtually embedded in a media (for example, the vibration of some game controllers for simulating shocks). Few works try to reconsider a given media [19] from a totally different perceptual point of view.

Empathic Media Transmission:

In cognitive science empathy is the recognition and understanding of the states of mind, beliefs, desires, and particularly, emotions of others. It is often characterized as the ability to "put oneself into another's shoes", or experiencing the outlook or emotions of another being within oneself; a sort of emotional resonance. For instance, if an observer sees another person who is sad and in response feels sad, that individual is experiencing empathy. Empathy can occur in response to cues of positive emotion as well as negative emotion. To qualify as empathy, the empathizer must recognize, at least on some level, that the emotion she or he is experiencing is a reflection of the other's emotional, psychological, or physical state. In addition to widen the sensory bandwidth, it is necessary to develop empathic media transmission, able to embed emotional cues in sensory-based coding of perceived events. The encoding is done directly into the physical expression of these additional senses by using suitable actuators integrated in a wearable, non-intrusive device. The audience who receives this multi-sensory information through the device and is thus in a state of partial sensorimotor immersion, will decode it for inducing emotions that ought to be identical to the emotions the creator of the media content intended to transmit.

Thanks to the synesthetic property of the newly defined media, and the empathic transmission of emotions that are contained in media, a more emotional link between the media and the audience is created. This is something already achieved for example in cinema through the background music or visual clues. The goal here is to improve the empathic relation to the media. How much this empathic link can be reinforced without breaking the duality audience / media is an interesting subject going beyond the scope of this presentation.

SENSES TO EMOTIONS

Following our definition of sympathetic media as a combination of synesthetic media and empathic transmission the translation / encoding of virtual emotional information into real / physical sensory information transmitted to the audience through actuators must be addressed. Two ways are proposed.

The first way to adress the problem of inducing emotions into the audience, is by extending the classical approach. This is done by adding sensory channels to the already present ones, usually sound and image. The relation of senses to emotions is studied for determining the most efficient ways to induce emotions from multi-sensory content. This includes the study of the attainable richness that a given sense or combination of senses can provide to encode virtual emotional content embedded in media content.

The second way to achieve the transmission of emotions is based on the enactive view and is thus favored. The method is to induce the emotional cause from its physiological consequences as perceived by the experiencing person. There is some evidence that this afferent feedback can modulate emotions (this is at the basis of the somatic-marker hypothesis [4] as well as the facial feedback hypothesis [1]). For example, a person experiencing stress or shame might have the feeling of a rise of the temperature. In the right context, rising effectively the temperature might help to induce the intended emotions, here stress or shame. Another technique is to use actuators to divert attention or generate subtle changes of emotional disposition [3]. Techniques such as surveys might help for this study by determining the best sensory channels and types of signals to use for inducing given emotions, with in mind works in cognitive science (enactive view), psychology and physiology.

To better see the difference in these two approaches, that are not exclusive, a second example is proposed. An emotion like sadness could be induced through visual (in a movie, dark atmosphere, rain, faces of the actors, etc...) and auditory (use of a certain type of music) clues. This is the classical approach. Sadness might be induced too by, for example, lowering the temperature and exercising slight pressures at appropriate locations on the body of the audience. While it can be argue that the first approach is already doing a good job at transmitting emotions, even without widening the sensory bandwidth; the second approach might be used in case of the absence of given sensory channels (for example, a radio program), the absence of a right context (looking at a movie on a portable device) or for audience with sensory disabilities.

SOFT AND HARDWARE FOR SYMPATHETIC MEDIATION

Existing media can be manually annotated or the emotions being automatically extracted. Given the difficulty to automatically extract emotional content from a given media, especially in the case of real-time applications, manual encoding will the first step in the creation of sympathetic media. Emotional tags could be considered to annotate the media in a way quite similar to the subtitles tracks on a DVD. An encoding module must be developed that encodes emotions to senses thanks to sets of rules and algorithms.

The hardware can be separated into two elements. One that supports the processing unit (notably the encoding module) and a transmitter and is interfaced with the media. This first element communicates with a second element that is a wearable device constituted of a receiver and the actuators. For this hardware part, we propose to design and conceive a wearable [22], unobtrusive, non-invasive, multi-actuators device, that will bring sympathetic media into homes in a similar way new technologies have brought cinema into homes through the Home Cinema.

Assessment of available actuators that can serve our purpose of providing relevant and efficient physical sensations and of being integrated in a wearable device will be conducted. As a first step, only actuators that act in non-invasive and external fashion relative to the human body will be considered. These actuators are, for example, actuators that can induce the following perceptual inputs: vibration, pressure, temperature, touch,... All these actuators act through the skin. Non-invasive actuators that act on internal organs of the human body (such as the galvanic vestibular simulation [20, 21]) will not be considered here but their existence will be discussed.

The design itself is another concern that can nonetheless be eluded at this stage. This device will contain a minimum of processing parts, except what is necessary for wirelessly communicating with the encoding module and for sending the received signals to the actuators. Because the actuators are non-invasive, the device which main function is to support these actuators will be non-invasive too. Nonetheless, most of the actuators are contact actuators and act through the skin. It implies that the device will be somehow attached to the body. For limiting the invasive feeling, the device will be for example designed as an armband. The future addition of other types of actuators will certainly lead to a reconsideration of the design, including location on the body.

FUTURE WORKS

Three types of future improvements can be foreseen.

- At the level of the senses in relation with the actuators. The progress in cognitive science and in nanotechnology make possible to think of new types of actuators that will be able to directly act on the brain of the audience without necessarily being invasive, and even directly induce emotions through electromagnetic signals [5], [2]. It should be noted that even in this case, encoding is necessary and that this type of brain stimulations can be somehow considered as a sense. Such actuators will inevitably rise ethical questions. At the same time, it opens the door to more immersive and affective virtual communication or experience.

- At the level of the emotions through their automatic extraction. The progresses in computing power, cognitive science, psychology, or semiotics makes us think that both the understanding of how emotions are induced and how to extract them automatically from media content will improve. The outcome of these advances will be useful to the future of this research.

- By implementing a mirror function to the whole system. The proposed system is aimed at transmitting emotions from a media to an audience. By adding sensors to the wearable device that could monitor the emotional state of the audience, a bi-directional empathic communication could take place with the possibility of interacting with the media. The media could ”react” to the emotional feedback of the audience.

ACKNOWLEDGEMENTS

The authors would like to thank Carson Reynolds for interesting insights and references.

REFERENCES

1. Buck, R. *Nonverbal behavior and the theory of emotion: the facial feedback hypothesis*. Journal of Personality and Social Psychology, 38, 811-824 (1980).

2. Padberg F. et al., *Prefrontal cortex modulation of mood and emotionally induced facial expressions : A transcranial magnetic stimulation study*, The Journal of neuropsychiatry and clinical neurosciences, vol. 13, no2, pp. 206-212 (2001)

3. C. Bassel and B. B. Schiff, *Unilateral vibrotactile stimulation induces emotional biases in cognition and performance*, Neuropsychologia, Volume 39, Issue 3, Pages 282-287 (2001)

4. Damasio, A.R. et al., *Somatic markers and the guidance of behaviour: theory and preliminary testing*, (pp. 217-229). H.S. Levin, H.M. Eisenberg A.L. Benton (Eds.). Frontal lobe function and dysfunction. New York: Oxford University Press, (1991)

5. Cook, CM. and Persinger, MA., *Experimental induction of the ”sensed presence” in normal subjects and an exceptional subject*. Percept Mot Skills. Oct;85(2):683-93, (1997)

6. A. Noe, *Action in perception*. 2004, Cambridge, MA: MIT Press.

7. E. Thompson, *Sensorimotor subjectivity and the enactive approach to experience*. Phenomenology and the Cognitive Sciences, 2005, 4: p. 407-427.

8. W. Prinz, *Perception and action planning*. European Journal of Cognitive Psychology. 1997, 9(2): pp. 129-154.

9. G. Knoblich and R. Flach, *Action identity: Evidence from self-recognition, prediction, and coordination.* Consciousness and Cognition, 2003, 12: pp. 620-632.

10. M. Wilson and G. Knoblich, *The case for motor involvement in perceiving conspecifics*. Psychological Bulletin, 2005, 131(3): pp. 460-473.

11. C. Baerveldt and P. Voestermans, *An enactive view on emotions*. 9th conference of the International Society for Theoretical Psychology (ISTP), June 3-8 2001, Calgary.

12. J. K. O’Regan and A. Noë, A *sensorimotor account of vision and visual consciousness*. Behavioral and Brain Sciences, 2001, 24: pp. 939-1031.

13. Antonio Damasio, *Descartes’s Error: Emotion, reason, and the Human Brain*. 1994, Avon Books.

14. G. Riva (Editor), F. Vatalaro (Editor), F. Davide (Editor) and M. Alcaniz (Editor), *Ambient Intelligence: The Evolution Of Technology, Communication And Cognition Towards The Future Of Human-Computer Interaction*. 2005, O C S L Press.

15. R. W. Picard, *Affective Computing*. 1997, MIT Press.

16. B. Myers, *A Brief History of Human Computer Interaction Technology*. ACM Interactions, 1998, 5(2): pp. 44-54.

17. T. Moon and G. J. Kim, *Design and evaluation of a wind display for virtual reality*. Proc. of the ACM symposium on Virtual reality software and technology, Hong Kong, 2004, pp. 122 . 128.

18. M. B. Khoudja et al., *Thermal Feedback for Virtual Reality*. International Symposium on Micromechatronics and Human Science, IEEE Conference, 2003, pp 153-158.

19. J. Bitton et al., *RAW: Conveying minimally-mediated impressions of everyday life with an audio-photographic tool*. Proceedings of CHI 2004 Conference on Human Factors in Computing Systems, 24 - 29 April 2004.

20. D. L. Wardman, et al., *Effects of galvanic vestibular stimulation on human posture and perception while standing*. J. Physiol., 2003, 551(3): pp. 1033-1042.

21. T. Maeda et al., *Shaking the World: Galvanic Vestibular Stimulation as a Novel Sensation Interface*, SIGGRAPH 2005.

22. T. Maeda et al., *Wearable Robotics as a Behavioral Interface -The Study of the Parasitic Humanoid-*. Proc. of the 6th International Symposium on Wearable Computers (ISWC.02), 2002.

Aural Antennae

Carson Reynolds
University of Tokyo
carson@k2.t.u-tokyo.ac.jp

Alvaro Cassinelli
University of Tokyo

Masatoshi Ishikawa
University of Tokyo

ABSTRACT

Aural Antennae are portable devices which translate sound impulses into vibrotactile stimulus. By swapping audio sensation for haptic sensation we illustrate one variety of artificial synesthesia. The compact devices can be worn to act as electronic travel aids for the hearing-impaired or used for augmented reality applications. Using a simple model of the audio scene's background noise, the device triggers when there is a large change in sound intensity from a specific direction.

Author Keywords

augmented reality, haptics, sensory substitution, hearing aids

ACM Classification Keywords

H.5.2 Haptic I/O
H.5.5 Sound and Music Computing
B.4.2 Input/Output Devices

ARTIFICIAL ANTENNAE

Suppose for a moment that your body was covered with several extremely long antennae. Like an insect, you use these antennae to probe about space, tapping and feeling the world that surrounds you.

For some, such a scenario is just a much-reduced plot of a Kafka story. However, we view this scenario in another light; our research group is preoccupied with how the precepts can be transformed to reproduce atypical experiences. We find motivation to create sensation similar to what the antenna-endowed insect feels.

Indeed, there are some surprising upshots to having antenna. It has been observed, for instance that cockroaches "use their antennae to detect a wall and maintain a constant distance" [2]. Antenna and cilia provide a variety of tactile spatial awareness. Some crude televised experiments with house cats and duct tape also show that felines use their hair to modify their gait and assess the space surrounding them [9].

Now suppose that you were covered with antennae which could pick up and localize minute aural signals. What would it be like to feel higher frequency audio signals in a manner to similar to how we already feel low-frequency bass?

UbiComp '08 Workshop W1 – Devices that Alter Perception (DAP 2008), September 21st, 2008.
This position paper is not an official publication of UbiComp '08.

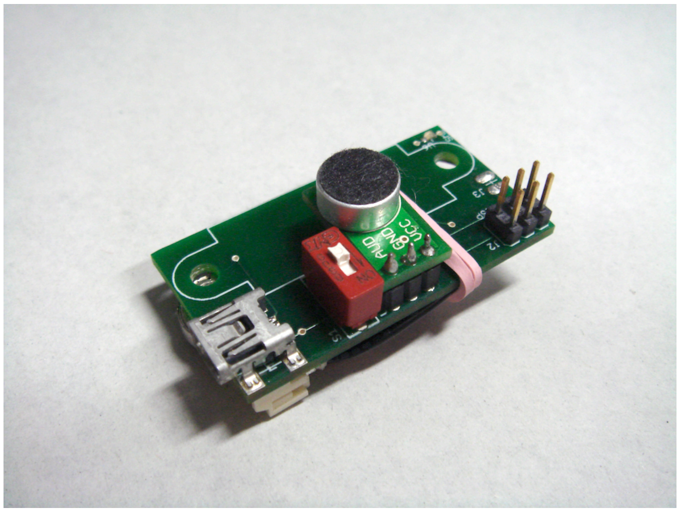


Figure 1. An Aural Antenna converts a signal from an electret microphone into vibrotactile stimulus.

HEARING IMPAIRMENT AND PROSTHESES

The Tadoma or Hofgaard method is a simple technique where those suffering from hearing loss feel the movements of a speaker by touching the parts of the face and neck used in speech production [17]. It has been used since the 1890s as a method for speech-reading [14].

As early as 1936, Gault discussed "hearing through the skin" and worked to develop mechanical apparatus for sound localization [7]. The development of vocoding techniques in the 1940s in turn spurred a variety of haptic audio systems modified to provide haptic stimulus at various loci on the body [17]. By the 1980s, wearable systems were constructed in which "speech sound generates a characteristic tactile pattern that observers can learn to identify" [20].

Wearable auditory systems gave way to implantables which were capable of "direct electrical activation of the auditory nerve" [24]. Further information about the neural basis of audition has been provided by studies of macaque monkeys using fMRI giving evidence integration of tactile and audio stimuli in the auditory cortex [11].

TRANSFORMATION OF PERCEPTION

Portable electro-mechanical systems make possible the creation of pattern converters or intermediaries that sit between our sense organs and the real world. The somatic nervous system, reflex arcs, and even muscles are organs whose artificial stimulation allows the transformation of perception.

That electrical activity has the ability to interact with the human percepts has been long known: "In his 1820 dissertation, Bohemian physiologist Johann Purkyne reported that a galvanic current flowing through the head upset balance and equilibrium" [6]. This technique has recently been employed by researchers who have built wearable devices to alter sense of balance as well as provide a "virtual sense of acceleration" [13].

Cutaneous rabbit illusion is an interesting perceptual illusion in which a series of taps produced by actuators at discrete locations feel as if they are interspersed between the actuators under particular timing conditions [8]. This phenomena has been exploited by a variety of haptic devices to provide stimulation in areas between actuators. For instance a 3 x 3 "rabbit" display composed of vibrator was used to communicate directional cues [22].

Another phenomena which has been exploited to transform perception is that of sensory substitution. Early attempts looked at using vibrating stimulators to convey visual pictures using an array built into a dental chair [1]. Experiments showed that visually impaired participants could "learn to recognize ... the layout of objects on a table in depth and in correct relationship."

Synesthesia (literally: joining of perception) has been induced in humans using a variety of methods, including electrical stimulation [5]. Less invasively, it may also be simulated through the use of devices which map the information of one senses onto another. This is the case with Fingersight devices, including one that allows wearers to feel optical edges as oscillations of a solenoid mounted above the fingertip [21].

We have developed a number of systems that seek to augment the percepts and specifically make use of the body or reflexes as part of interaction [18]. Earlier work on laser-based tracking systems [15] led us to think of how optical based information might be felt by users, which led us to radar and antennae as metaphors for interaction.

HAPTIC ANTENNAE

We began to experiment with the concept of artificial antennae as part of device illustrating another concept: Haptic Radar [4]. This is a project that seeks to augment spatial awareness by creating radar out of sensors which act to extend the range of touch for the skin.

As most humans have a copious amount of hair located on their head (at least at some point in their life), and our heads are something we wish to protect, we reasoned a headband device would be a good first form factor to test.

We devised a system linking pairs of infrared rangefinders to motor vibrators into a circular arrangement. An earlier paper, Augmenting spatial awareness with Haptic Radar, details experimental results concerning the Haptic Radar. Most saliently, we found that 86% of untrained participants could use the system to move to avoid objects they could not see [3].

Following these initial experiments, we began a redesign with the aim to make individual, compact, Haptic Antenna. To replace the Arduino board, we selected an ATMEL ATtiny13 RISC microcontroller for its compact size (4 mm x 4mm). The process of reading from infrared rangefinder and controlling a vibrating motor requires a minimum of computational resources so this 8-bit microcontroller operating at 20 MHz is adequate.

After recreating and testing the system on breadboard, we added a 100 milliampere-hour lithium-ion polymer battery as well as charging circuitry. After testing this through-hole technology circuit, we designed and fabricated a surface-mount technology printed circuit board (using the freely available Eagle printed circuit board CAD software.)

After further testing and circuit board revisions, we have arrived at a Haptic Antenna in a much more portable instantiation. The device melds a microcontroller, infrared rangefinder, motor-vibrator (a common part in portable phones), battery and electronics. Altogether, these components occupy 25 cm^3, which is a factor of 34 times smaller than the previous version's electronic system.

AURAL ANTENNAE

During this process we came to ask ourselves: what if people felt directional sound as opposed to distance information? Imagine that a car is honking behind you but that you cannot hear it because of a hearing impairment or environmental noise. Now imagine that the honking could be felt on the body at the location nearest to the car's horn.

As a starting point to test this concept we have been building prototype audio-to-touch sensory substitution devices. Aural Antennae are compact, worn modules which produce vibrotactile stimulus in response to audio signals emanating from a particular direction.

Principle of Operation

Our current prototype builds upon the precious Haptic Antennae platform. Instead of a range finder, we attach a daughter board containing an electret microphone, conditioning resistors and capacitors as well as an OPA344 operational amplifier configured with a gain of $G = 100$.

The analog voltage output of the amplifier is digitized using the ATtiny's internal 10 bit analog to digital converter. The microcontroller's firmware samples the microphone at approximately $f_s = 9000Hz$.

After each sample, the microcontroller computes a simple moving average (SMA) over the previous $k = 10$ samples (1). The absolute difference (δ) is then computed between the current sample s_t and SMA (2).

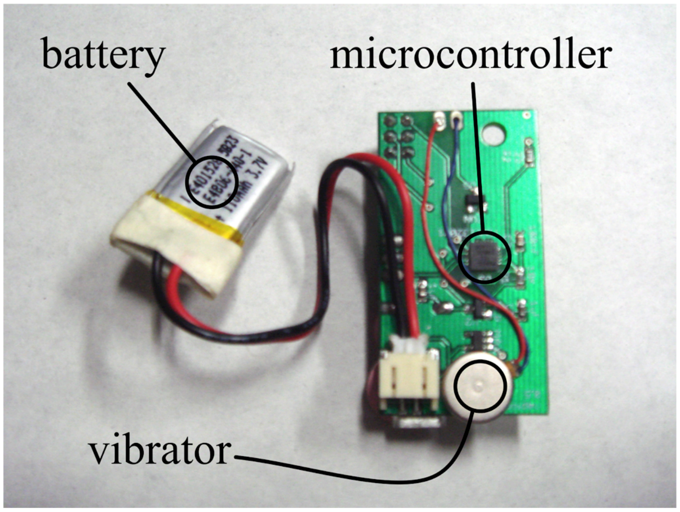


Figure 2. An Aural Antenna module incorporating lithium-ion polymer battery, 20 MHz, 8-bit microcontroller, and vibrotactile motor.

$$SMA = \frac{s_t + s_{t-1} + \cdots + s_{t-(k-1)}}{k} \quad (1)$$

$$\delta = |s_t - SMA| \quad (2)$$

If δ is greater than $\frac{2^{10}}{10}$ (10% of the dynamic range of the analog to digital converter), then the vibrator is activated with 100% duty cycle until the next sample is processed. This moving average works as an extremely rudimentary adaptive background noise filter. The vibrating motor is controlled by a MOSFET transistor whose gate is tied to a digital output pin of the ATtiny microcontroller.

Our initial experiments with Haptic Antennae indicated that blindfolded participants readily interpreted the vibrotactile stimulus and associate it with approaching objects. We expect that similar phenomena will be observed in forthcoming experiments with the aural antennae.

The device exploits our innate ability to process (in a parallel manner) haptic stimulus applied to skin or the Vellus hair which covers most areas of our bodies. Other recent work on electronic travel aids [16] as well as the use of vibrotactile cuing in virtual environments [12] make use of this phenomena. Experiments have also documented that strong haptic stimulus can induce a startle reflex [25], which may be useful in emergency situations.

EXTENSIONS

While independent modules may be worn simultaneously, when networked together the augmentations provided by the devices would be greatly enhanced. We are in the process of evaluating low-power wireless chips such as Zigbee to incorporate into the modules. We anticipate that wireless antennae would be able to work together to provide "rabbit" perceptual illusions of motion between the actuators.

Making use of shotgun-type microphones has improved the directionality of our initial prototype. The use of laser-microphones might increase range significantly. With network capabilities we could create a worn antenna array capable of sound localization using time-of-arrival.

One can imagine a type of wearable simultaneous localization and mapping (SLAM) system. This could be a fusion of antenna-array sound localization and laser ranging and detection (LADAR). Such a system might use a Bayesian network to estimate object location based on data provided by both audio and optical sensing systems.

Another extension of this work is in the area of actuation. The "pancake" style vibration motor we are using (KOTL C1030B028F) has the advantage of being compact, but presents substantial initial friction which makes response somewhat limited. Other researchers have reported on the use of air puffs and acoustic cues to elicit startles [23]. Still other researchers have thoroughly investigated using electrical stimulation to provide haptic cues [10].

AS OTHER SPECIES HEAR

We have developed an example of aural antennae which provide haptic feedback. Often thinking about haptic devices is constrained by our experience of our existing senses. We have instead sought to break with this convention by seeking to emulate insect perception.

Thinking more openly, we can imagine a myriad of new biomimetic ways of seeing the world. Compound eyes and ocellus suggest worn garments that have thousands of cameras. Mimicry of insect's abilities to acutely detect subtle vibrations [19] and act on this information could lead to extension of touch in the manner that optics have extended the sight.

ACKNOWLEDGMENTS

The authors would like to thank Tomohiko Hayakawa, Kenichiro Otani, and Alexis Zerroug for work on early prototypes.

REFERENCES

1. P. Bach-Y-Rita, C. C. Collins, F. A. Saunders, B. White, and L. Scadden. Vision substitution by tactile image projection. *Nature*, 221(5184):963–964, March 1969.
2. J. M. Camhi and E. N. Johnson. High-frequency steering maneuvers mediated by tactile cues: antennal wall-following in the cockroach. *J Exp Biol*, 202(Pt 5):631–643, March 1999.
3. A. Cassinelli, C. Reynolds, and M. Ishikawa. Augmenting spatial awareness with haptic radar. In *Wearable Computers, 2006 10th IEEE International Symposium on*, pages 61–64, 2006.
4. A. Cassinelli, C. Reynolds, and M. Ishikawa. Haptic radar. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Sketches*, New York, NY, USA, 2006. ACM.

5. R. Cytowic. Synesthesia: Phenomenology and neuropsychology. *Psyche*, 2(10):2–10, 1995.
6. R. C. Fitzpatrick and B. L. Day. Probing the human vestibular system with galvanic stimulation. *J Appl Physiol*, 96(6):2301–2316, June 2004.
7. R. H. Gault. Recent developments in vibro-tactile research. *Journal of the Franklin Institute*, 221(6):703–719, June 1936.
8. F. A. Geldard and C. E. Sherrick. The cutaneous "rabbit": a perceptual illusion. *Science*, 178(57):178–179, October 1972.
9. Japan's greatest mysteries: gaffer tape, April 14th 1996. http://www.glumbert.com/media/cattape.
10. H. Kajimoto, N. Kawakami, T. Maeda, and S. Tachi. Tactile feeling display using functional electrical stimulation. In *In Proceedings of the 9th International Conference on Artificial Reality and Telexistence*, 1999.
11. C. Kayser, C. I. Petkov, M. Augath, and N. K. Logothetis. Integration of touch and sound in auditory cortex. *Neuron*, 48(2):373–384, October 2005.
12. R. W. Lindeman, J. L. Sibert, E. Mendez-Mendez, S. Patil, and D. Phifer. Effectiveness of directional vibrotactile cuing on a building-clearing task. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 271–280, New York, NY, USA, 2005. ACM Press.
13. T. Maeda, H. Ando, T. Amemiya, N. Nagaya, M. Sugimoto, and M. Inami. Shaking the world: galvanic vestibular stimulation as a novel sensation interface. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Emerging technologies*, New York, NY, USA, 2005. ACM.
14. S. J. Norton, M. C. Schultz, C. M. Reed, L. D. Braida, N. I. Durlach, W. M. Rabinowitz, and C. Chomsky. Analytic study of the tadoma method: Background and preliminary results. *J Speech Hear Res*, 20(3):574–595, September 1977.
15. S. Perrin, A. Cassinelli, and M. Ishikawa. Laser-Based Finger Tracking System Suitable for MOEMS Integration. *Proceedings of Image and Vision Computing, New Zealand (IVCNZ)*, pages 131–136, 2003.
16. S. Ram and J. Sharf. The people sensor: A mobility aid for the visually impaired. *iswc*, 00, 1998.
17. C. M. Reed, N. I. Durlach, and L. A. Delhorne. Historical overview of tactile aid research. In *Proceedings of the second international conference on tactile aids, hearing aids and Cochlear Implants*, 1992.
18. C. Reynolds, A. Cassinelli, and M. Ishikawa. Meta-perception: reflexes and bodies as part of the interface. In *CHI '08: CHI '08 extended abstracts on Human factors in computing systems*, pages 3669–3674, New York, NY, USA, 2008. ACM.
19. D. Robert and M. C. Göpfert. Novel schemes for hearing and orientation in insects. *Current Opinion in Neurobiology*, 12(6):715–720, December 2002.
20. F. Saunders, W. Hill, and B. Franklin. A wearable tactile sensory aid for profoundly deaf children. *Journal of Medical Systems*, 5(4):265–270, December 1981.
21. G. Stetten, R. Klatzky, B. Nichol, J. Galeotti, K. Rockot, K. Zawrotny, D. Weiser, N. Sendgikoski, S. Horvath, and S. Horvath. Fingersight: Fingertip visual haptic sensing and control. In *Haptic, Audio and Visual Environments and Games, 2007. HAVE 2007. IEEE International Workshop on*, pages 80–83, 2007.
22. H. Tan and A. Pentland. Tactual displays for wearable computing. *Personal and Ubiquitous Computing*, 1(4):225–230, December 1997.
23. B. K. Taylor, R. Casto, and M. P. Printz. Dissociation of tactile and acoustic components in air puff startle. *Physiology & Behavior*, 49(3):527–532, March 1991.
24. B. S. Wilson, C. C. Finley, D. T. Lawson, R. D. Wolford, D. K. Eddington, and W. M. Rabinowitz. Better speech recognition with cochlear implants. *Nature*, 352(6332):236–238, July 1991.
25. J. S. Yeomans, L. Li, B. W. Scott, and P. W. Frankland. Tactile, acoustic and vestibular systems sum to elicit the startle reflex. *Neuroscience & Biobehavioral Reviews*, 26(1):1–11, January 2002.

Learn Traffic State Based on Cooperative Localization

Zhuo Sun, Yiqiang Chen, and Juan Qi
Pervasive Computing Center, Institute of Computing Technology
Chinese Academy of Sciences, Beijing, China
{sunzhuo, yqchen, qijuan}@ict.ac.cn

ABSTRACT

A key problem of monitoring the traffic state is the localization and tracking of vehicles and passengers. In this paper, we present a new cooperative localization technique which makes use of collaboration of mobile phones for traffic monitoring. Instead of relying on the signal strength only, our cooperative localization approach utilizes additional connection information among mobile phones to improve the localization accuracy. It separates mobile phones into different clusters by some short-range links such as Bluetooth, then locates the members in a cluster simultaneously. We design a simulation experiment and the results show that our method is able to catch the main trace of all the members in a cluster. In this way, the traffic flows can be monitored.

Author Keywords

Traffic monitoring, Cooperative localization, Wi-Fi networks

ACM Classification Keywords

H.5.3 [Information Interfaces and Presentation (e.g., HCI)]: Group and Organization Interfaces - Collaborative computing

INTRODUCTION

In modern society, the demand for traffic monitoring systems that can detect position, velocity, density, and flow rate in a street is increasing. Traditional monitoring approaches include GPS, video monitoring etc. However, GPS as a standalone system is not sufficient for an increasing number of transport applications due to limited visibility to the satellites within the urban canyons. Video monitoring is able to record the whole traffic state, but it takes large human efforts to analyze the traffic data. Since the prevalence of mobile phones, cellular networks are once under consideration for user localization. But the low localization accuracy of a few hundred meters makes it not practical in real world.

Recent years, Wi-Fi is popular around the world. Many cities carry out 'Wi-Fi City' projects to implement large-scale coverage, especially in the important streets. Wi-Fi based localization gets an accuracy of about 20 meters outside. As more and more mobile phones are equipped with wireless cards that can receive Wi-Fi signals, it is possible to monitor the traffic state by Wi-Fi networks. A mobile phone receives Wi-Fi signals from several access points (APs) and the signal strength implies the distance from the mobile phone to APs. There are various methods to determine the mobile phone location according to the signal strength, which we will discuss later. However, due to the noisy and fluctuant characteristics of Wi-Fi signals, the location estimation by a single mobile phone may be rather biased. Although some filters such as [3] are advised for smoothing, the results seem not apparently improved in some bad situations.

In this paper, we propose a novel cooperative localization in Wi-Fi based networks, which makes use of collaboration of mobile phones to learn the traffic state. As far as we know, this concept is completely new in Wi-Fi based localization domain. In the traditional non-cooperative scenario, a mobile phone infers the location only from its own received Wi-Fi signals. While in a cooperative scenario, several mobile phones nearby are combined in a cluster by some short-range links to determine locations simultaneously. For example, Bluetooth, which is commonly used on mobile phones can be used to detect the other devices in a short range (usually $10 \sim 20$ meters). The devices within the valid range is considered as a cluster. The mobile phones in the same cluster can exchange data about their traces. Although each single trace may be inaccurate, it is more likely to find a reliable trace of the cluster by integrating the trace information together. Figure 1 illuminates such a scenario. In addition to signal strength, connectivity information or distance measurements among phones in a same cluster are utilized to improve localization accuracy. One problem with the cluster is that the cluster members are dynamic due to their mobilities and it has to be updated periodically.

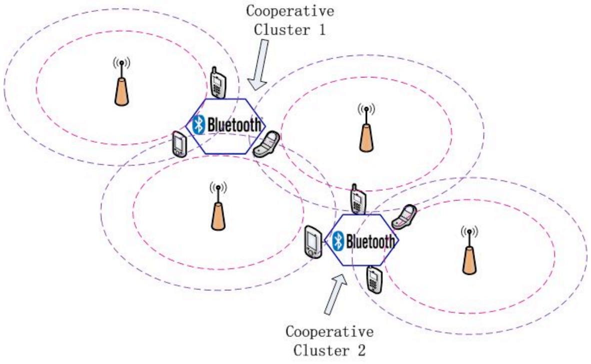


Figure 1. The cooperative localization scenario in Wi-Fi networks.

UbiComp '08 Workshop W1 – Devices that Alter Perception (DAP 2008), September 21st, 2008.
This position paper is not an official publication of UbiComp '08.

RELATED WORK

Let us consider the localization problem in the similar fields such as robots and wireless sensor networks. The robots also interact with the environment and each other through imperfect sensor measurements corrupted by noise. Consequently, the concept of cooperative localization is introduced, where groups of robots combine sensor measurements to implement cooperative localization. [8] demonstrates the utility of introducing a second robot to aid in the tracking of the exploratory robot's position.

In wireless sensor networks, cooperative localization uses connectivity information (such as who is within the communication range of whom or estimated distances between neighbors) to derive the locations of the nodes in the network. [10] first builds a relative map through a multidimensional scaling, then with three or more anchor nodes, the relative map can be transformed and absolute coordinates of all the nodes are computed.

[5] proposes a cooperative positioning technique by utilizing the additional information obtained from short-range links like WiMAX/Wi-Fi to enhance the location estimation accuracy in cellular networks. However, it just fuses these two kinds of signals together and its cooperative positioning refers to fusion of different sensor data.

WI-FI BASED LOCALIZATION METHODS

In general, Wi-Fi based localization research can be classified into two main categories: deterministic techniques and probabilistic techniques.

Deterministic techniques [1], [6], [2] use deterministic inference methods to estimate a user's location. The RADAR system developed by Microsoft Research [1] proposes nearest-neighbor heuristics and triangulation methods to infer a user's location. It maintains a radio map which tabulates the signal strength received from different access points at selected locations. Each signal-strength measurement is then compared against the radio map and the coordinates of the best matches are averaged to give the location estimation. The accuracy of RADAR is about three meters with 50 percent probability. The LANDMARC system [6] exploits the idea of reference points to alleviate the effects caused by the fluctuation of RFID signal strength. The accuracy is roughly one to three meters. However, the placement of reference tags should be carefully designed since it has a significant effect on the performance of the system. Moveover, the RFID readers are so expensive that it is infeasible for localization in a large area. In [2], an online procedure based on feedback from users was employed to correct the location estimation of the system.

Another branch of research is the probabilistic techniques [9], [11], [4] which construct a conditional probability distribution over locations in the environment of interest. In [4], Ladd et al. use probabilistic inference methods for localization. They first use Bayesian inference to compute the conditional probability over locations, based on received signal-strength measurements from nine access points in the environment. Then, a postprocessing step, which utilizes the spatial constraints of a user's movement trajectories, is used to refine the location estimation and reject the results with significant change in the location space. Depending on whether the postprocessing step is used or not, the accuracy of this method is 83 or 77 percent within 1.5 meters. In addition, Roos et al. [9] compare the performance of the non probabilistic nearest-neighbor method with that of two probabilistic approaches. The results show that the two probabilistic approaches produce better results than the nearest-neighbor method and the average location estimation error is below two meters. Furthermore, the time-series analysis technique [11] was introduced to study the correlation among consecutive samples received from the same access point over time. The authors reported that better accuracy can be achieved by taking such correlation into account.

Since Hidden Markov Model (HMM) can utilize both single samples and user trajectories in the form of sequential knowledge [4], we adopt it to solve the tracking problems. HMM is used to model the user traces by treating physical locations as hidden states and the signal strength measurements as observations. Each user trace is denoted as $T = \{l_1, l_2, ..., l_n\}$ and $l_i = (x_i, y_i)$ is considered as a discrete physical location with x and y coordinates. $O = \{o_1, o_2, ..., o_n\}$ is defined as observation space and $o_j = \{s_1, s_2, ..., s_m)$ is a set of signal strength measurements from m different access points. In this way an HMM can be defined as a quintuple (T, O, λ, A, π), where λ is prior probability of signal distribution. T, O, λ can be learned from some collected traces. Transition matrix A describes how a person travels through the state space, which is constrained by physical locations and mobile velocity. π is an initial location-state distribution encoding where the user initially may be. Generally, it is set by a uniform distribution if there is no prior knowledge about user locations. In the online localization phase, given an observed signal sequence $\widetilde{O} = \{\widetilde{o_1}, \widetilde{o_2}, ..., \widetilde{o_n}\}$, the well-known Viterbi algorithm [7] can be used to infer the most probable physical space sequence $\widetilde{T}$.

COOPERATIVE LOCALIZATION

In a cooperative scenario, not a single but a group of mobile phones determines their locations together. The primary discrimination of the group is that each member is close to others and has a similar direction of movement. We call this group a *Cooperative Cluster*, as shown in Fig.1. The members in the same cluster communicate and exchange data with each other. Then the integrated information can be used to improve respective localization results since more constraints are added. In traffic flow monitoring, overall trends is focused on rather than single trace of each vehicle. While in a large-scale area, the distance between nearby mobile phones can be ignored. That's to say, each member in a cluster shares a similar trace, which represents most of the members' trends. We will mainly discuss the solution to this problem in the following part.

The discrimination of cooperative clusters depends on links that have shorter range such as Bluetooth, UWB, Zigbee.

Mobile phones within the range for a while are considered to have similar behaviors and can form a cooperative cluster. We prefer Bluetooth because its prevalence on mobile phones. Bluetooth is a standard and communications protocol primarily designed for low power consumption with a short range in each device. It enables these devices to communicate with each other when they are in range. Maximum permitted power of about $2.5mW$ can cover an area of circle with a radius of about 10 meters. Therefore, we use Bluetooth to detect whether the mobile phones are in a same cooperative cluster or not. The request to keep in touch with each other for a while guarantees the same direction of movement, and it does not make sense if mobile phones pass by toward the contrary direction.

After cooperative clusters have been divided, we predict the direction of each cluster using the integrated information gathered from members of the cluster. Suppose there are n members in a cluster in a time period t_1 to t_p, indicated by $\{M_1, M_2, ..., M_n\}$. We apply the HMM method introduced ahead on each member M_i separately to get a trace $T_i = \{l_1^{(i)}, l_2^{(i)}, ..., l_p^{(i)}\}, i = 1, 2, ..., n$. We propose a Trace Least Square (TLS) algorithm to find the optimized trace $T_c = \{l_1^{(c)}, l_2^{(c)}, ..., l_p^{(c)}\}$ of the cluster. TLS addresses to the following optimization problem:

$$\arg\min_{T_c} \sum_{i=1}^{n}\sum_{j=1}^{p}(l_j^{(c)} - l_j^{(i)})^2 + \sum_{j=2}^{p}(l_j^{(c)} - l_{j-1}^{(c)})^2 \qquad (1)$$

The first term penalizes the discrepancy between T_c and T_i, and the second term keeps T_c as smooth as possible.

However, this optimization is unsolvable due to p parameters in all. Consider the trace information is time dependable, we break it into single time period and it is equivalent to:

$$\begin{cases} \arg\min_{l_j^{(c)}} \sum_{i=1}^{n}(l_j^{(c)} - l_j^{(i)})^2 & \text{if } j = 1 \\ \arg\min_{l_j^{(c)}} \sum_{i=1}^{n}(l_j^{(c)} - l_j^{(i)})^2 + (l_j^{(c)} - l_{j-1}^{(c)})^2 & \text{if } j > 1 \end{cases} \qquad (2)$$

It can be explained that the location at t_1 is decided by all $l_1^i, i = 1, 2, ..., n$, and the location at $t_j, 1 < j \leq p$ is depended on not only all $l_j^i, i = 1, 2, ..., n$ but also the location at t_{j-1}. Therefore, the optimization is easily solved and we get:

$$l_j^{(c)} = \begin{cases} \frac{1}{n}\sum_{i=1}^{n} l_j^{(i)} & \text{if } j = 1 \\ \frac{1}{n+1}(\sum_{i=1}^{n} l_j^{(i)} + l_{j-1}^{(c)}) & \text{if } j > 1 \end{cases} \qquad (3)$$

All the members in a cluster can be represented by the trace T_c. The work flow of the algorithm is prescribed in Fig.2. As mobile phones are not static, the clusters need to be updated periodically. Therefore, the overall algorithm is in a circulation.

SIMULATION EXPERIMENTS

To evaluate the performance of cooperative localization, we make a simulation experiment in an indoor environment. A Wi-Fi wireless environment is established on the 3rd floor of our academic building, with an area of about $30m \times 15m$. The whole layout of the test-bed is shown in Fig.3, and 5 TENDA APs are deployed around. We choose the hallway to simulate a street, and two persons take different mobile phones to represent the passengers. One mobile phone is an O2 Xda Atom Life smartphone and the other one is a Nokia N95, both with Bluetooth and Wi-Fi wireless cards.

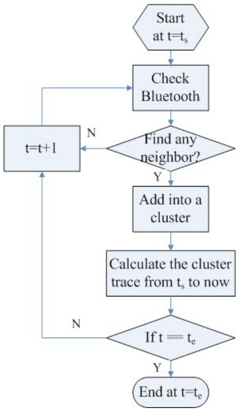


Figure 2. The work flow of cooperative localization.

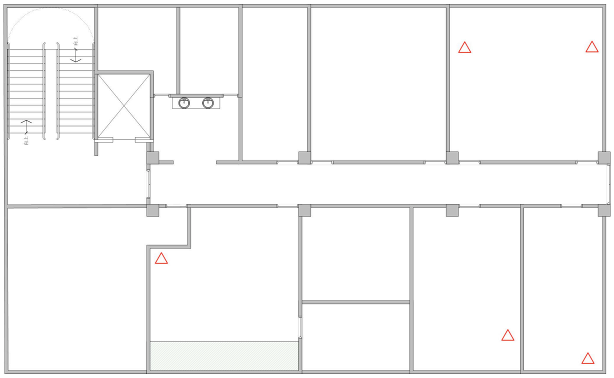
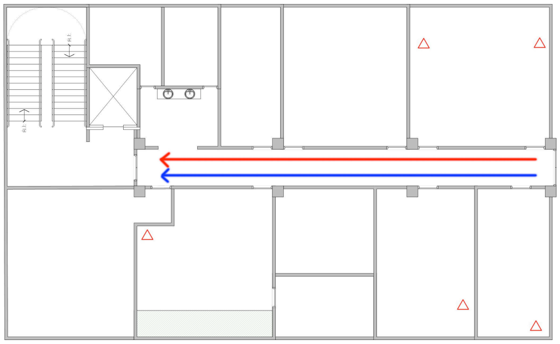


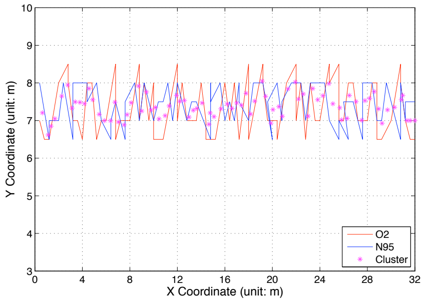
Figure 3. The whole layout of the test-bed.

We design two different traces to simulate the traffic state. The first one is that these two persons go straight forward side by side, as shown in Fig.4(a). The red line denotes the actual trace of the person with O2, and the blue one is with N95. As they are always in the Bluetooth range, they are considered to be in a cluster. We apply our cooperative localization to track each one separately and the cluster also. The tracking traces and the cluster trace are illustrated in Fig.4(b). It is obvious that the tracking traces got by one mobile phone separately are quite fluctuant, although we have given geographic constraints in the HMM models. However, the cluster trace reflects a more smooth trace that represents the overall trend of members. It can be inferred that our cooperative localization can catch a group of traffic objects that have similar behaviors.

The second trace we design is one person (with N95) turns into a room halfway as shown in Fig.5(a). After they get

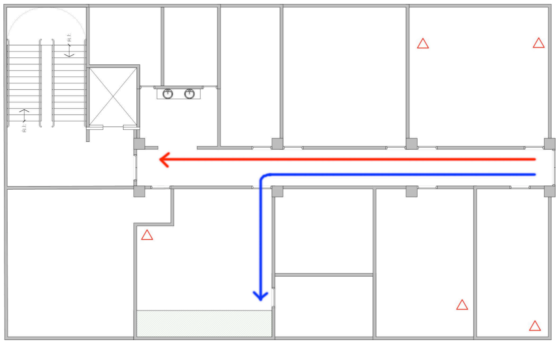


(a) Actual traces.

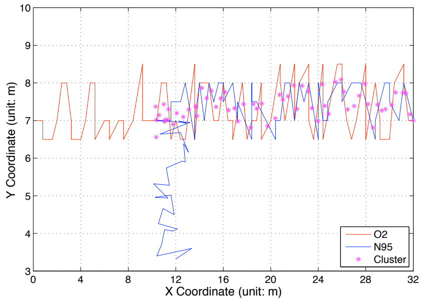


(b) Tracking traces and cluster trace.

Figure 4. Two persons go straight forward side by side.



(a) Actual traces.



(b) Tracking traces and cluster trace.

Figure 5. One person turns apart halfway.

apart from each other, the Bluetooth signals are blocked by the wall and these two mobile phones can not get in touch again. Figure 5(b) presents the tracking result of this trace. Each single trace is still rough due to the noisy signals while the cluster trace is relatively smooth. Moreover, as the two phones exceed the range of Bluetooth, the cluster is broken up and the cluster trace stops at the turning point accordingly. It indicates that our cooperative localization can divide the mobile phones into exact clusters, make sure the members in a cluster have similar mobile trends.

CONCLUSION AND FUTURE WORK

In this paper, we originally propose a new cooperative localization technique in Wi-Fi based networks to learn traffic state. Unlike the cooperative localization in robots and wireless sensor networks, our cooperative localization makes use of collaboration of mobile phones to separate them into different clusters. Then their behaviors can be represented by the cluster trace. We simulate it in wireless environment and the simulation results show the effectiveness. In the future, we will make actual traffic experiments and improvement on the algorithms further.

REFERENCES

1. P. Bahl, A. Balachandran, and V. Padmanabhan. Enhancements to the radar user location and tracking system. 2000.
2. E. S. Bhasker, S. W. Brown, and W. G. Griswold. Employing user feedback for fast, accurate, low-maintenance geolocationing. In *PERCOM '04: Proceedings of the Second IEEE International Conference on Pervasive Computing and Communications (PerCom'04)*, page 111, Washington, DC, USA, 2004. IEEE Computer Society.
3. F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P. Nordlund. Particle filters for positioning, navigation, and tracking. 2002.
4. A. M. Ladd, K. E. Bekris, A. Rudys, L. E. Kavraki, and D. S. Wallach. Robotics-based location sensing using wireless Ethernet. *Wireless Networks*, 11(1), 2005.
5. C. L. F. Mayorga, F. D. Rosa, S. A. W. G. Simone, M. C. N. Raynal, J. Figueiras, and S. Frattasi. Cooperative positioning techniques for mobile localization in 4g cellular networks. In *Proceedings of IEEE International Conference on Pervasive Services (ICPS'07)*, 2007.
6. L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil. Landmarc: Indoor location sensing using active rfid. In *PERCOM '03: Proceedings of the First IEEE International Conference on Pervasive Computing and Communications*, page 407, Washington, DC, USA, 2003. IEEE Computer Society.
7. L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. pages 267–296, 1990.
8. I. Rekleitis, G. Dudek, and E. Milios. Multi-robot collaboration for robust exploration. *Annals of Mathematics and Artificial Intelligence*, 31(1-4):7–40, 2001.
9. T. Roos, P. Myllymaki, H. Tirri, P. Misikangas, , and J. Sievanen. A probabilistic approach to wlan user location estimation. In *Intl J. Wireless Information Networks, vol. 9, no. 3, pp. 155-164, July*, 2002.
10. S. Y, R. W, and Z. Y. Localization from mere connectivity. In *Proceedings of the fourth ACM international symposium on Mobile ad hoc networking and computing (MOBIHOC 2003)*, Annapolis,MD,USA, 2003.
11. M. Youssef and A. Agrawala. The horus wlan location determination system. In *MobiSys '05: Proceedings of the 3rd international conference on Mobile systems, applications, and services*, pages 205–218, New York, NY, USA, 2005. ACM.

Spatial coverage vs. sensorial fidelity in VR

Alexis Zerroug, Alvaro Cassinelli, Masatoshi Ishikawa
University of Tokyo
alvaro@k2.t.u-tokyo.ac.jp

ABSTRACT

With this paper we wish to promote a discussion about the different forms that an immersive VR system can take. This will be done by reflecting on a controversial concept: that of a totally immersive (understood as multimodal) but partial (understood as limited to a part of the body) virtual reality interface. The proposed concept of total/partial immersiveness may be seen as a new orthogonal dimension in the taxonomic classification of systems in the 'virtuality continuum' introduced in [2]. An interesting aspect of the proposed configuration is the possibility for it to be wearable. We will briefly describe the motivation for this new taxonomic dimension from a theoretical point of view, as well as explain the practical reasons that lead us to this concept. This will be done by discussing earlier work from one of the authors that illustrates the possibilities of a total immersive VR system but also pinpoints a number of inescapable limitations.

Author Keywords

virtual reality, immersive system, multi-modal, haptic

ACM Classification Keywords

H.5.1 Multimedia Information Systems — Artificial, augmented, and virtual realities,
H.5.2 User Interfaces — Haptic I/O

INTRODUCTION

Since the sixties and pioneered by the Sensorama simulator (a multimodal system created by Morton Heilig in 1962 [3]), lots of immersive systems were developed with different technologies and goals in mind. The main driving force was perhaps the entertainment industry with its clear goal of immersing the user as much as possible in a simulated environment governed by laws and rules of a specific gameplay. In this context, total immersion could be contemplated as the Holy Grail of Virtual Reality since it would afford the gamer to forget for a moment the (physical or social) constraints of the real world. However, researchers on the emerging field of Virtual Reality kept innovating with other goals in mind such as developing systems for training, learning, medical therapy and data visualization. Anticipating the development of highly immersion-capable technology, it appeared relevant to answer the question of how much immersion was going to be really necessary to succeed in each and all of these goals. Unsurprisingly, it turns out that the answer is extremely dependent on each particular goal. Moreover, it soon became clear that the sense of subjective *presence* in a virtual environment does not necessarily account for the level of objective sensorial *immersion* [4]. It may even be the case that immersion in an almost (but not completely) perfect simulation would provoke the user to distantiate himself from the rendered environment (there may be an 'uncanny valley' [1] for artificial *reality* as a whole, not just with respect to realistic humanoid robots). In an effort to clarify the relation between immersion and presence in the virtual environment, as well as related concepts such as coherent spatial perception and realistic interaction, some authors developed a taxonomy of virtual reality systems [2] which is useful as it introduces the concept of a 'virtual continuum' spanning the realm of the completely real to the completely virtual world, and qualifiy whatever is in between these extremes as 'mixed reality'.


Figure 1. photograph of the 'Time Machine'

We would like to discuss in this workshop the possibility of a total immersive interface (that is, reproducing with high fidelity most basic sensorial modalities and therefore belonging somehow to the totally virtual), whose action is restricted to a part of the body (and therefore making it impossible to classify it as an interface completely rendering a virtual environment). There has been a lot of research on enhancing a Head Mounted Display with binaural sound and other kinds of actuators; in a way, such a device would be the archetype of a total/partial immersive system, but we would like to discuss the possibility of deploying such configuration to other

UbiComp '08 Workshop W1 – Devices that Alter Perception (DAP 2008), September 21st, 2008.
This position paper is not an official publication of UbiComp '08.

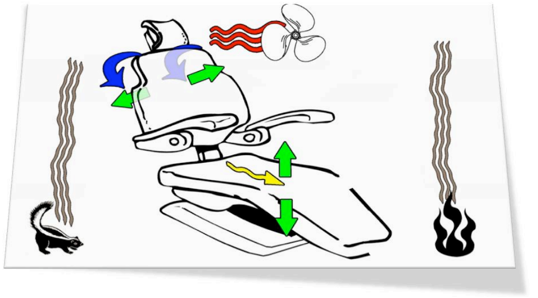


Figure 2. Schematic picture of the actuators

parts of the body (not necessarily encompassing all the sensorial organs). An example would be a box with an aperture for introducing a limb: when the user does so, he will experience as if his arm or leg is in another environment, say, a space filled with water and sea creatures he can touch and feel as real. Of course, one can argue that body propioception, vestibular sensation and visuospatial input are all basic senses that a 'total immersion interface' should be able to reproduce, and that this is in direct contradiction with the idea of a partial interaction with the body. However, in certain cases this is an apparent limitation: compelling presence within the simulated environment may be effective with this sort of interface - even though not all the sensorial apparatus is engaged, in particular thanks to task-oriented (bottom-down) perceptual blindness [5]. Of course, there has been some research on interfaces capitalizing on the limitations of human attentional span or the physical limitations inherent to the visual organs (the best known being the foveal displays [5]). We are however interested in another issue: that of an artificially generated *sense of presence of a part of the body* within a virtual environment. Evidence for compelling partial presence (i.e. partial body relocation) is described in [6]. Lastly, a practical motivation for the proposed concept is its compatibility with a wearable realization.

TOTAL IMMERSIVE SYSTEMS

A total immersive system needs to deal with at least two fundamental problems: the first is how to properly generate artificial sensory stimuli; the other is how to avoid the stimuli from the real world to interfere with the simulation (i.e. achieving sensory deprivation). Futuristic brain-computer interfaces may achieve both goals at once (c.f the 'neural plug' in the movie the Matrix, described earlier by pioneering writer William Gibson in his 1984 novel Neuromancer). Present day more or less invasive BCI enable elementary motor control [7] or generate sensations that would overlap with the external world stimuli if these are present [8], [9]. More conventional systems such as the CAVE [10] or HMD-based VR systems may instead capitalize on real world stimuli in order to enhance the realism of the immersiveness, but this is done at the expenses of the freedom of the simulation (i.e., one must constrain certain aspects of the simulated world such as the orientation on space, gravity and ground texture). With respect to the CAVE, the HMD-based configuration enables a limited form of body sensory deprivation - perhaps by immersing the rest of the body in a liquid or making the user relax on a bed or chair. The latter approach has been tried in an earlier experiment by one of the authors [11]. The intent of the experiment was to create a realistic sense of presence in the virtual world (a WWI battlefield), while at the same time cutting the subject from the real world sensory input. But can we imagine a system capable of totally immersing a part of his body in another world, while still capable of creating a (partial) sense of presence and sufficient emotional arousal?

Figure 3. redering of the virtual environment

EARLIER WORK

'Time Machine: VERDUN 1916' [11] is an immersive system build by one of the authors that 'sends' users back in time at the site of Verdun (a battlefield during World War I). The system achieves a high level of immersion thanks to a HMD and number of different actuators described in the following (figure 2).

A commercial stereoscopic HMD (the Z800 3DVisor from eMagin with a diagonal FOV of approx. 40 deg) and inertial head tracking was used to render the simulated environment (figure 3). Thanks to the information provided by the inertial sensors, the user was able to look around while tied on a modified dentist chair. The chair could tilt and vibrate as a whole (to simulate explosions) then providing some form of vestibular stimulation, and was also covered with dozens of tactile actuators to simulate the ground texture (as the wounded avatar was being dragged on the floor). A belt covering the torso was fit with sixteen vibrators and was used to render the footsteps of a rat walking over the lying body (figure 4).

The HMD is fitted with noise cancelling ear bud speakers, but a pair of large isolating headphones seemed more efficient in reducing interferences from the real environment. Additional speakers and a subwoofer were used to render low frequency sounds produced by the shock waves of virtual explosions. Since air flow can greatly enhance the feeling of presence on an open (virtual) space, a fan was installed to simulate wind as well as heat waves. Finally, an air-pump connected to a box containing chemicals (figure 5) would bring the smell of powder and dead corpses.

Figure 4. photograph of the vibrators inside the chair and the belt

The Time Machine was exhibited at the 2007 Laval Virtual international conference on VR. During a five day long exhibition, more than 300 people tried this immersive experience. Nobody was indifferent and some people were disoriented for a couple of minutes when 'coming back to the reality'. Also, two individuals asked to stop the 'Time Machine' because they grew scared. But of course, the machine was not conceived to function as a ghost train in a fun fair: there was no gratuitous surprise, nor rendering of blood or explicit scenes of fighting. One can wonder what aspect of the experience was more scary for these people: the emotionnaly charged context (i.e. the simulated battlefield), or the fact that they were immersed in a realistic, multimodal VR environment for the first time in their lives. From the technical point of view, this experiment demonstrates that low-cost immersive systems are not dreams anymore; also, it shows that the combination of a relatively low number of discrete multi-modal actuators is enough to create a completely immersive experience and make forget the low view angle of the HMD for example.

The team received two Awards at Laval Virtual; the best prize of the Student competition and the best prize of the IVRC Jury with an invitation to participate to the final step of the International Virtual Reality Contest in Japan.

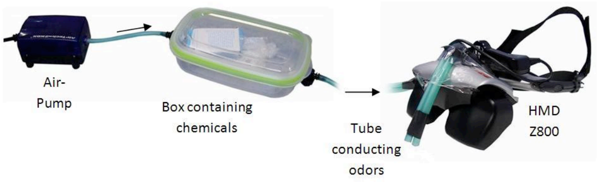


Figure 5. photograph of the odor system and the HMD

DISCUSSION

A subject having tried the system says: "I've never been immersed in such a way. I've never been emotionally immersed, and that's really an incredible experience. That's exactly the kind of immersion I was waiting for in a virtual world. There was no interactivity apart from the head movement, but maybe thats the reason why it worked" [12].

This comment is enlightening: the experience is believable precisely because the story being simulated matches the limitations of the interface: the subject is a wounded soldier, and as such *cannot move*. There may be many cases when a proper design can get around the limitations of the interface (in this case, it's inability to arbitrarily generate artificial propioceptive stimulation); however, one problem faced with the Verdun simulator was its bulkiness, immobility as well as the necessity of one or more technical operators for a unique subject in the machine.

CONCLUSION

There is some evidence that realistic auditory and haptic stimuli might be more important than realistic visuals when treating some types of phobia using VR systems [13]. This means that a total/partial immersive system not involving the sense of vision may be able to accommodate this type of simulation. An example would be for instance a wearable globe extending on the forearm that would create the impression of walking spiders and/or the temperature of virtual bodies. An early prototype of such a device is described in figure 6 [14]. From the point of view of the taxonomy described in [2], the total/partial immersive system can be seen as the counterpart of the 'window-on-the-world' mixed reality systems (these are monitor based, non-immersive video displays showing real scenes upon which computer generated imagery is electronically overlaid). Indeed, the proposed configuration can be seen as a window on the *virtual* world, not necessarily encompassing the visual senses, but instead the rest of the perceptual modalities. Perhaps a better analogy would be that of a spatio-temporal wormhole or a portal to another world. It is partial in the sense of it being a window located at a specific place in (real) space where the user can introduce a part of the body. As said before, this makes compatible the

Figure 6. photograph of the 'Ants glove' [14]

notion of partial/total immersion with that of a wearable interface, as opposed with total immersive systems where the user is completely immersed in the virtual world. An second potential advantage of such system could be that if there is an uncanny valley for artificial environments, as suggested in the introduction, then it may constitute an advantage that these systems secure a cognitive distance between the rendered environment and the user's 'reality'.

ACKNOWLEDGEMENT

We would like to thank all the people that contributed to the construction of the 'Time Machine': Jrmy Hautreux, Pierre Le Gargasson, Benot Malabry and Antoine Lelarge, Clment Mona, Julien Pg, as well as their professors M. Crison and M. LeRenard and M. Geslin. Thanks to Carson Reynolds for interesting suggestions and feedback.

REFERENCES

1. Ho, Chin-Chang and Macdorman and Karl F. and Pramono, Dwi Z. A. D. Human emotion and the uncanny valley: a GLM, MDS, and Isomap analysis of robot video ratings In *HRI '08: Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction, Amsterdam, The Netherlands*, (2008), 169–176

2. P. Milgram and F. Kishino. A Taxonomy of Mixed Reality Visual Displays. In *IEICE Transactions on Information Systems*, volume E77-D (1994).

3. M. L. Heilig. Sensorama simulator. Patent 3050870 (1962).

4. D. A. Bowman and R. P. Mcmahan. Virtual Reality: How Much Immersion Is Enough?. In *Computer*, volume 40 (2007), 36–43.

5. K. Cater and A. Chalmers and P. Ledda. Selective quality rendering by exploiting human inattentional blindness: looking but not seeing. In *VRST '02: Proceedings of the ACM symposium on Virtual reality software and technology*, ACM (2002), 17–24.

6. M. Botvinick and J. Cohen. Rubber hands 'feel' touch that eyes see.. In *Nature*, volume 391 (1998).

7. L. R. Hochberg and M. D. Serruya and G. M. Friehs and J. A. Mukand and M. Saleh and A. H. Caplan and A. Branner and D. Chen and R. D. Penn and J. P. Donoghue. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. In *Nature* volume 442, Nature Publishing Group, 164–171.

8. W. H. Dobelle. Artificial vision for the blind by connecting a television camera to the visual cortex. In *ASAIO journal*, American Society for Artificial Internal Organs (2000), 3–9.

9. T. Maeda and H. Ando and M. Sugimoto. Virtual Acceleration with Galvanic Vestibular Stimulation in Virtual Reality Environment. In *VR '05: Proceedings of the 2005 IEEE Conference 2005 on Virtual Reality*, IEEE Computer Society (2005), 289–290.

10. C. Cruz-Neira and D. J. Sandin and T. A. DeFanti and R. V. Kenyon and J. C. Hart. The CAVE: audio visual experience automatic virtual environment. In *Commun. ACM*, ACM (1992), 64–72.

11. C. Mona and A. Zerroug and J. Hautreux and P. Le Gargasson and B. Malabry and J. Pégé and A. Lelarge. Time Machine: VERDUN 1916. http://www.time-machine.info/, ESIEA / ESCIN (2007).

12. S. Kuntz. A VR Geek Blog / Laval Virtual 2007. http://cb.nowan.net/blog/2007/04/23/laval-virtual-2007/

13. A. S. Carlin and H. G. Hoffman and S. Weghorst. Virtual reality and tactile augmentation in the treatment of spider phobia: a case report.. In *Behaviour research and therapy*, volume 35 (1997), 153–158.

14. K. Sato and Y. Sato and M. Sato and S. Fukushima and Y. Okano and K. Matsuo and S. Ooshima and Y. Kojima and R. Matsue and S. Nakata and Y. Hashimoto and H. Kajimoto. Ants in the Pants HOW?. http://www.mushi-how.com/, The University of Electro-Communications (2008).

15. S. Bouchard and S. Côté and J. St-Jacques and G. Robillard and P. Renaud. Effectiveness of virtual reality exposure in the treatment of arachnophobia using 3D games. In *Technol. Health Care*, IOS Press (2006), 19–27.