ENACTIVE08

5th International Conference on Enactive Interfaces

19-21 November, 2008, Scuola Superiore Sant'Anna, Pisa, Italy

Edited by:

Emanuele Ruffaldi

Marco Fontana

Preface

Enactive Interfaces and Systems are a new generation of Human- Computer Interfaces (HCI) that are based on the concept of Enactive knowledge, that is the knowledge acquired by doing. These new interfaces allow to make old applications more intelligent and responsive and to create new kinds of applications.

The introduction of the Enaction concept in HCI has created a multidisciplinary research community capable of integrating theoretical model with interaction paradigms and advanced technologies like visualization systems, haptics and spatialized audio.

The Enactive Conference is an important annual meeting occasion for researchers in the field of Enactive Interfaces. This conference series has been started by in 2004 by the European Network of Excellence.

On the behalf of the Organization Committee we would like to welcome you to the proceedings of ENACTIVE08, the fifth edition of the International Conference on Enactive Interfaces. ENACTIVE08 is held in Pisa at Scuola Superiore Sant'Anna on 19th-21st November 2008.

We would like to take this opportunity to thank to all the authors and the reviewers of the papers for their precious contribution and commitment. Furthermore we would like to thank all the people of the Committees for their efforts and the Scuola Superiore Sant'Anna for the support.

Emanuele Ruffaldi
Marco Fontana

PERCRO Laboratory - Scuola Superiore S.Anna

Organization

General Chairs

Emanuele Ruffaldi and Marco Fontana
PERCRO - Scuola Superiore S.Anna

Program Chairs

Technology and Applications

Marcelo Wanderley
McGill Univrsity

Emilio Sánchez
CEIT

Human Action and Perception

Elena Pasquinelli
COMPAS

Olivier Gapenne
UTC

Tutorials and Demo Chairs

Gualtiero Volpe
DIST - University of Genova

Franco Tecchia
PERCRO - Scuola Superiore S.Anna

Scientific Committee

Massimo Bergamasco
PERCRO - Scuola Superiore S.Anna

Benoit Bardy
University of Montpelier 1

Nadia Magnenat-Thalmann
MIRALAB - University of Geneve

Roberto Casati
Institut Nicod

Marcelo Wanderley
McGill University

Sile O'Modrian
Queen University of Belfast

Yasuyoshi Yokokoji
Kyoto University

Carlo Alberto Avizzano
PERCRO - Scuola Superiore S.Anna

Antonio Frisoli
PERCRO - Scuola Superiore S.Anna

Award Chairs

Benoit Bardy
University of Montpelier 1

Aude Billard
EPFL

Hisato Kobayashi
Hosei University

Local Committee

Massimo Bergamasco
PERCRO - Scuola Superiore S.Anna

Franco Tecchia
PERCRO - Scuola Superiore S.Anna

Marcello Carrozzino
PERCRO - Scuola Superiore S.Anna

Fiammetta Ghedini
PERCRO - Scuola Superiore S.Anna

Vittorio Spina
PERCRO - Scuola Superiore S.Anna

Haakon Faste
PERCRO - Scuola Superiore S.Anna

Elisabetta Sani
PERCRO - Scuola Superiore S.Anna

Davide Vercelli
PERCRO - Scuola Superiore S.Anna

Program Committee

Malika Auvray
Federico Avanzini
Carlo Alberto Avizzano
Federico Barbagli
Benoit Bardy
Delphine Bernardin
Frederic Bevilacqua
Diego Borro
Alexandre Bouënard

Marcello Carrozzino
Torquato Cecchini
Bhaskar Chakraborty
Hanne De Jaegher
Gunnar Declerck
Haakon Faste
Manuel Ferre
Jean-Loup Florens
Antonio Frisoli
Ilja Frissen
Olivier Gapenne
Hugues Genevois
Jorge Juan Gil
Bruno Giordano
Teresa Gutiérrez
Sylvain Hanneton
Vincent Hayward
Pablo F. Hoffmann
Gunnar Jansson
Armen Khatchatourov
Charles Lenay
Jean-Claude Lepecq
Charlotte Magnusson
Joseph Malloch
Daniel Mellier
Javier Melo
Daniel Mestre
Emma Murphy
Sandra Nogry
Elena Pasquinelli
Olivier Penelaud
Otniel Portillo Rodriguez
Carsten Preusche
Nicolas Rasamimanana
Arcadio Reyes-Lecuona
Agnès Roby-Brami
Jorge Rodríguez
Victor Rosenthal
Katia Rovira
Emilio Sánchez
Oscar Osvaldo Sandoval-Gonzalez
Joan Savall
Basilio Sierra
Stephen Sinclair
Regis Thouvarecq
Paolo Tripicchio
Luis Unzueta
Gualtiero Volpe
Marcelo Wanderley
Sabine Webel

Contents

Technology

Online Passive Reconstruction -Based Bilateral Control under Time-Varying Communication Delay and Packet-Loss

Ali Hosseini* Jordi Artigas† Anja Brunberg‡ Carsten Preusche† Gerd Hirzinger†

(*) Engineering Science Department, Simon Fraser University, Canada
(†) Institute of Robotics and Mechatronics, DLR–German Aerospace Center, Germany
(‡) Institute of Automatic Control–IRT, RWTH Aachen University, Germany
E-mail: aha44@sfu.ca, jordi.artigas@dlr.de, a.brunberg@irt.rwth-aachen.de, carsten.preusche@dlr.de, gerd.hirzinger@dlr.de

Abstract

In force reflecting teleoperation systems, master and slave mechanisms are connected through a communication means to exchange information. Most communication channels induce time delays which make the control design process difficult from the stability point of view. The wave variable method introduced in the mid-90's, based on the theory of passivity, ensures the passivity of the communication block in presence of arbitrary constant time delays.

This paper demonstrates that in channels with time-varying delays and data-loss (typical for the Internet and satellite radio communication links), the passivity of the channel using the basic wave variable method cannot be guaranteed. Modifications are applied to this method to cope with the effect of data-loss and variations of delay to provide the passivity framework for stability proof. The method is implemented on two 7 degrees of freedom DLR–LWR robots. The test results are presented and show the effectiveness of our method.

1. Introduction

Force reflecting teleoperation describes systems involving two distant, yet coupled robots where a human operator moves the local robot, also known as master, and the measured motion commands are transmitted to a distant located robot, also known as slave, which tries to track the commands of the master device. In this system the operator is coupled to the remote environment through the electromechanical intermediate elements. In some tasks, where the direct presence of the human operator is not possible (due to harsh environments, geometrical limitations or distance issues), it is desirable to have a slave manipulator in the remote site, commanded by the human operator from a distant location. Examples of this sort can be found in nuclear plants for handling toxic materials (harsh and health threatening environment), on-orbit satellite maintenance (distance, risk and cost issues), or laparoscopic keyhole surgery (invasiveness, surgeon posture and healing period issues). In these applications, the haptic perception of the interaction between the robotic device and the manipulated object by the user is essential in order to successfully complete the required operation. Adding the haptic feelings to the hand of the remote operator means that the forces and torques themselves will induce dynamics to the human movements. The slave system plus the environment on the remote site and the master system and the human movements on the local site interact dynamically. This type of teleaction is called *bilateral telemanipulation*.

Haptic displays are designed to serve as master devices and generate the feeling of direct manipulation through replication of the interaction forces/torques generated at the slave/environment site. Proper control design has to provide safety for the human operator as well as high performance to fulfill the given task and provide a high level of telepresence feeling at the same time. In addition to the stability, which is an essential requirement for control systems, high levels of transparency are sought. Transparency in its ideal form means that the human operator feels the same forces and motions as if he/she is directly manipulating the environment. Review on the different control schemes used in force reflecting telemanipulation can be found in [8, 12, 6]. Lawrence in [10] introduced the general four channel (4CH) haptic control architecture which is theoretically capable of providing perfect transparency.

It was shown there that a trade-off between transparency and stability requirements always exists. The problem is intensified when the communication between the master and slave is delayed. A stable telemanipulation system in an undelayed scenario might become unstable in an operation over a communication line with a small fraction of a second time delay.

One of the approaches toward the problem induced by time delay is using the passivity concept. Stability proof can be drawn from providing the passivity condition for teleoperation system sub-blocks. This method provides an intuitive and robust means for designing the delayed teleoperation systems. This concept was introduced by Anderson and Spong [1] and Niemeyer and Slotine [14]. The basic wave variable method guarantees the passivity of the communication channel and hence ensures overall stability over the channels with arbitrary constant time delays.

For channels with time-varying delays, modifications over the basic wave variable method were introduced by different authors. Munir and Book used a modified smith predictor, a kalman filter and an energy regulator to improve the performance of wave based teleoperation for the time-varying delayed channels like the Internet [13]. The effect of time varying delay has been studied by Yokokohji et al. [16]. Lozano [11] showed the conditions of the time-varying delay function which can lead to communication channel activity induced unstability, using continuous time mathematics. Chopra et al. [5] adapted the earlier work of Lozano et al. to discrete time and designed a method to handle the time-varying issue in a stable way with a high tracking precision. Hirche and Buss studied the effect of packet-loss and data reordering and designed a class of reconstruction algorithms to provide the passivity of the communication channel [7]. Berestesky et al. [4] introduced an algorithm for passive reconstruction of data and buffering in bilateral teleoperation over the Internet.

In this paper a novel data processing and reconstruction method based on online energy observation and packet processing during the teleoperation over the Internet or satellite communication is introduced and developed.

2. Background

2.1. A Wave-based Teleoperation Scheme

Wave variables present a modification or extension to the theory of passivity which provides robustness to arbitrary (unknown) constant time-delays and is based on the concepts of power and energy. The method is applicable to nonlinear systems with contact to unknown and unmodeled environments [14]. (u, v) are introduced as the new transformed variables, called the wave

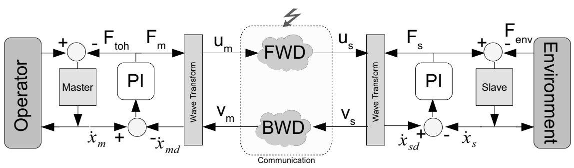


Figure 1: A wave-based PEFF teleoperation control scheme.

variables. Their definition is based on a pair of power variables $(\dot{x}, F)$ or any other mechanical power conjugates. u is the right moving wave which is transmitted from master to the slave and v is left moving wave transmitted from slave to the master. In Fig.1 a traditional wave based Position-Exchange/Force-Feedback (PEFF) bilateral control scheme is depicted. Two PI controllers at master and slave sites are responsible for minimizing the end effector position errors between the master and slave mechanisms, hence generating the required manipulation force on slave site and haptic interaction forces to the operator's hand via master mechanism. Before and after the communication channel, wave transformers are implemented, encoding and decoding the wave signals to energy domain (force/velocity) signals and vice versa. The wave transformation equations for this scheme are as (1) for the master site and (2) for the slave site.

$$u_m(t) = \frac{b\dot{x}_{md}(t) + F_m(t)}{\sqrt{2b}},\ v_m(t) = \frac{b\dot{x}_{md} - F_m}{\sqrt{2b}}. \tag{1}$$

$$u_s(t) = \frac{b\dot{x}_{sd}(t) + F_s(t)}{\sqrt{2b}},\ v_s(t) = \frac{b\dot{x}_{sd}(t) - F_s(t)}{\sqrt{2b}}. \tag{2}$$

b is the wave impedance constant, F_m and F_s are local controller force commands and $\dot{x}_{md}$ and $\dot{x}_{sd}$ are desired motion velocities of the master and slave respectively.

2.2. Passivity Condition

Defining P_{in}, the power entering a system, as the scalar product between input vector x and output vector y, and E_{store} as the stored energy of the system, the system is passive if and only if

$$\int_0^t P_{in} \mathrm{d}\tau = \int_0^t x^T y \mathrm{d}\tau \geq E_{store}(t) - E_{store}(0). \tag{3}$$

The power input to the communication channel of a teleoperation system for each time instance is calculated as

$$P_{in}(t) = \dot{x}_m(t)F_m(t) - \dot{x}_s(t)F_s(t). \tag{4}$$

The following inequality has to be conserved for a communication channel to act as a passive system:

$$\int_0^t P_{in}(\tau)\mathrm{d}\tau = \int_0^t [\dot{x}_m F_m - \dot{x}_s F_s]\mathrm{d}\tau \geq E_{store}(t) - E_{store}(0). \quad (5)$$

Reformulating the energy balance equation from the power domain to the wave variable domain, the equation will be as

$$\int_0^t \frac{1}{2}[u_m^2(\tau) + v_s^2(\tau) - u_s^2(\tau) - v_m^2(\tau)]\mathrm{d}\tau \geq E_{store}(t) - E_{store}(0). \quad (6)$$

Without loss of generality we can assume that the initial stored energy in the channel is equal to zero, $E_{store}(0) = 0$. We also assume that the forward and backward delays are constant. Then the shift of wave signal can be formulated as (7).

$$u_m(t - T_{fwd}) = u_s(t), \quad v_m(t) = v_s(t - T_{bwd}). \quad (7)$$

Combining (7) and (6) with few mathematical manipulations, the new passivity condition is obtained as

$$\frac{1}{2}\left[\int_{t-T_{fwd}}^t u_m^2(\tau)\mathrm{d}\tau + \int_{t-T_{bwd}}^t v_s^2(\tau)\mathrm{d}\tau\right] \geq 0, \quad (8)$$

which holds as far as equation (7) is true (constant time delay and guarantee of signal delivery). So the stability of the teleoperator system is guaranteed.

2.3. Problem Statement

In teleoperation over computer networks and space radio links, the problem of time delay and data delivery should be taken into account. Communication protocols exist that guarantee the delivery of the information and maintain the signal orders (e.g. TCP), however, they induce extra time delays and it is shown in [13] that they are not suitable for teleoperation. UDP protocol is used normally for network-based teleoperation due to its faster exchange rate (the smaller the delay, the better the operation performance and transparency) but this protocol is prone to packet-drops and reordering because the transmission is not controlled in network layers.

2.3.1 Time-Varying Delay

Let us consider that the forward and backward time delays are not constant and are defined by time functions $T_{fwd}(t)$ and $T_{bwd}(t)$. Contrary to the assumption in [11], equation (7) should be changed into the following (yet ignoring the packet-loss):

$$u_m(t) = u_s(t + T_{fwd}(t)), \; v_m(t + T_{bwd}(t)) = v_s(t).$$

The reason is that $T_{fwd}(t)$ and $T_{bwd}(t)$ functions are unknown at time t to the master and slave sent waves $u_m(t)$, $v_s(t)$ respectively. These functions can be identified only after transmission and delivery. Therefore the passivity condition in form of equation (8) cannot be obtained. However the passivity condition can be conserved by keeping the decoupled forward and backward channel lines passive, as the following:

$$E_{comm,fwd} = \frac{1}{2}\int_0^t [u_m^2(\tau) - u_s^2(\tau)]\mathrm{d}\tau \geq 0, \quad (9)$$

$$E_{comm,bwd} = \frac{1}{2}\int_0^t [v_s^2(\tau) - v_m^2(\tau)]\mathrm{d}\tau \geq 0. \quad (10)$$

2.3.2 Packet-Loss

Using the Internet as the communication medium will induce occasional packet drops and information loss in exchange line, in addition to jitter and variations of delay. The same is valid for satellite based communication where the time delay function is smoother and the data loss rate is lower. Both mediums are discrete time systems where senders and receivers sample and pick the data with a certain frequency. An increase in time delay leads to the *empty sampling instances* for the receiver which is called signal stretching in continuous domain. The decrease in time delay leads to instances where more than one data exists to be sampled at the receiver site and is called signal compression in continuous domain. However only one signal can be sampled and the other will be discarded, leading to information loss. It is important to notice that the wave information has energetic meaning. Another source of packet-loss is the reordering of the sent signals where a later sent packet arrives earlier than an earlier sent packet. For empty sampling instances (occurring due to previous reasons) two straightforward approaches can be used. Either replacing the empty instance with the Zero value (Null packet or Zeroing strategy) or conveying the last valid signal also known as Hold-last-sample (HLS). It has been shown before that none of them is sufficient for performance or stability conditions and modification should be applied over them [7]. The Null packet strategy is lossy, overly conservative and leads to poor performance as well as wear and noise in mechanical parts. On the other hand, HLS strategy keeps the transmitted wave form well, but cannot guarantee the passivity of the channel and may lead to instability of the operation [9].

3. Proposed Method

For the sake of simplicity we go on with neglecting the loss and time delay variations of the backward channel and just consider the forward channel issues (dealing with backward link as ideal). After reformulating the continuous time expression in (9) for discrete time, we obtain the relation below:

$$2\Delta E(i)/T_s = \sum_{k=0}^{k=i} u_m^2(k) - \sum_{k=0}^{k=i} u_s^2(k), \qquad (11)$$

where $\sum_{k=0}^{k=i} u_m^2(k)$ is the sum of the square of the right moving wave variables and represents the power input to the forward communication block, $\sum_{k=0}^{k=i} u_s^2(k)$ is the sum of the squared received waves which is the power output of the communication and T_s is the sampling time. We calculate the overall sent energy and put it in a data structure where after transmission over the time delayed communication link it will arrive at the remote site (here slave). So the remote receiver at time t will be informed about the overall energy sent up to a certain time instance (t^*). $t = nT_s$ is the current time and $t^* = n^*T_s$ is the time stamp of the arrived packet. These two times are related by equation (12).

$$nT_s = n^*T_s + T_{fwd}(n^*T_s) \qquad (12)$$

To ensure the passivity, it is enough to keep the following expressions in (13).

$$E_{sent}(n^*T_s) - E_{received}(nT_s) \geq 0 \qquad (13)$$

We name equation (13) as the *online forward observed energy* equation. The passivity of the channel can be checked comparing the overall energy input and output at the same time instances. However, in practice the passivity condition cannot be measured online. Due to accumulative property of the sum of the sent wave energies, the relation (14) holds

$$E_{sent}(n) \geq E_{sent}(n^*) \text{ where } n > n^* . \qquad (14)$$

Considering the forward channel and renaming E_{sent} by E_m and $E_{received}$ by E_s, and based on the argument in (14), the following relation is valid

$$E_m(n) - E_s(n) \geq E_m(n^*) - E_s(n). \qquad (15)$$

The left hand side of (15) is the passivity condition and the right hand side is the online forward observed energy equation. Keeping the inequality in (13) always valid, leads to fulfilling the following overall passivity condition.

$$E_m(n) - E_s(n) \geq 0 \qquad (16)$$

We use the online forward observed energy equation for online passivity checks. As it is apparent, using this equation for passivity observation gives the system safety limits against activity and eventual unstability but it is more conservative than what is actually needed in reality. A packet generator is implemented at each sender site putting the algorithm's required information in a data structure. The information put into the sending packets are sending time stamp (n^*T_s), current sending wave value ($u_m(n^*)$), packet reception acknowledgment flag, summation of the overall sent wave ($T_s \sum_0^{n^*} u_m(n^*T_s)$) and the wave energy sent overall up to the current instance ($\frac{1}{2}T_s \sum_0^{n^*} u_m^2(n^*T_s)$). Without getting into programming details which can be found in [9], our program performs the following tasks to ensure the passivity. At each sampling time the packet reader on the receiver side checks the flag for arrival of data. When the flag indicates data arrival and its time stamp is bigger than the last recorded time stamp the packet data will be processed. This process consists of one-step-ahead energetic checks and if the online energy observer violated the passivity condition, the current wave is modified to dissipate required energy and keep the passivity. The program checks what would happen energetically if the current wave was conveyed using equation (17). E_{curr} is the energy that the latest delivered wave could transmit to the system during one sample time.

$$E_{curr}(n) = \frac{1}{2}u_m^2(n^*)T_s \qquad (17)$$

The forward energy observer (E_{feo}) checks the safety of conveying this current arrived wave in relation (18). $E_{s_{output}}(n-1)$ is the energy output of the packet processor up to current time which is fedback to the algorithm and $E_{sent}(n^*)$ is the sum of the energy sent from the master up to n^* instance.

$$E_{feo}(n) = E_{sent}(n^*) - E_{s_{output}}(n-1) - E_{curr}(n) \qquad (18)$$

When (18) is bigger or equal to zero, passivity would be kept by conveying $u_m(n^*)$. If (18) is negative, $u_m(n^*)$ should be modified and the arrived $u_m(t^*)$ would be changed in a way to dissipate the activity of $E_{feo}(n)$. If $\hat{u}_s^2(n) \geq 0$ then a real answer exists and with wave modification in one sample the sensed activity can be dissipated. We can refer to $u_m(n^*)$ as $u_s(n)$. The modification is applied on the $u_s(n)$ based on the following energy balance equation in (19).

$$\frac{1}{2}u_s^2(n)T_s + E_{feo}(n) = \frac{1}{2}\hat{u}_s^2(n)T_s \qquad (19)$$

Based on the energy balance equation, $u_s(n)$ has to be changed to $\hat{u}_s(n)$ following:

$$\hat{u}_s^2(n) = u_s^2(n) + \frac{2E_{feo}(n)}{T_s} \qquad (20)$$

to dissipate the $E_{feo}(n)$ activity. The sign of the wave (positive or negative) can be interpreted as push or pull command. Even though wrong signature selection does not theoretically affect the energetic behavior of channel it is important to be decided correctly. This is more significant when a black-out occurs in the line and high number of consequent losses are detected. The trend of the sent command (push or pull) is observable by comparing the currently arrived sum of the sent waves with the last valid recorded (LVR) arrived sum in (21). n^{lvr} is the last-valid-recorded time instance of data arrival and is smaller than n^*.

$$S = \sum_{i=0}^{i=n^*} u_m(i) - \sum_{i=0}^{i=n^{lvr}} u_m(i) \tag{21}$$

When S is positive the trend indicates more push commands and when it is negative indicates pull. So the conveyed signal after a black out will be as (22).

$$\hat{u}_s(n) = \frac{S}{|S|}\sqrt{u_s^2(n) + \frac{2E_{feo}(n)}{T_s}} \tag{22}$$

For simple empty instances (due to reordering, occasional loss or stretching) selection of the signature, simply based on the current wave is correct but equation (22) generalizes these conditions as well.

When $\hat{u}_s^2(n) < 0$ the wave is replaced with a Zero which is the maximum energy that can be dissipated in one instance. The remaining undissipated energy is recorded to be dissipated in the next samples *if necessary*. As the activity detection in forward observed energy equation is conservative and preemptive and is based on the worst possible case (sending zero command from counterpart), correction of this undissipated excess energy (remained after Zeroing) might not be even necessary in the next steps. However it is decided by the algorithm based on the received new data from the sender.

When the time stamp of the arrived packet is older than the last valid recorded time, the packet is discarded (The algorithm has already dealt with this packet's absence energetically as another later-stamped-earlier-arrived packet informed the receiver of the energy sent in between). Furthermore, there can be instances with no packet delivery. In both cases the algorithm has to handle the empty instance. The algorithm assumes that the current wave should be equal to the last valid signal (HLS strategy). With this assumption the HLS signal should be checked in the previous algorithm for energy considerations. If needed the same modifications introduced before will be applied on the waves. The scheme of the proposed method and how it is integrated to the forward link is depicted in Fig.2.

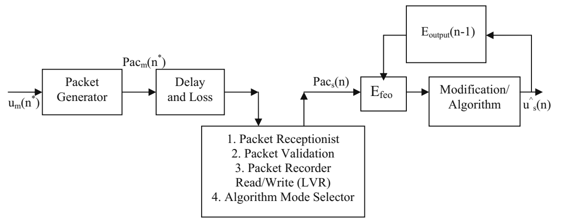


Figure 2: Scheme of the forward link packet processing and energy compensation.

4. Simulation

A wave based communication channel is simulated in Simulink and MATLAB using Sfunction. The channel is capable of emulating different time delay patterns and different levels of information loss. In Fig. 3 a sinusoidal wave signal is fed to the channel. The time delay function is a ramp as it is shown in subplot (e) and the packet-loss rate is 50% (an extreme high loss condition). Sampling rate is 1kHz. Presence of the increasing time-delay and feeding waves with decreasing absolute value and using HLS startegy would result in the worst case scenario for passivity of the channel. Our designed simulation test provides a tough trial condition to prove the potential of the method. The energetic behavior of the HLS versus our compensation method can be seen in subplot (c). The energy dissipation and wave modification effects can be seen in subplot (a). The rectangular box shows the Zeroing (maximum energy absorption) and the circular box shows wave modifications.

In this scenario the HLS strategy cannot hold the passivity and the energy curve goes to the negative portion but our compensation method keeps the passivity (Fig. 3c). Observing Fig. 3a and Fig. 3c, one can see that whenever the E_{feo} function goes to the negative part a compensation action is performed on the waves and results in keeping the overall passivity condition.

5. Experiment

The robots used in our tests are DLR-LWR-III and DLR-LWR-II as master and slave mechanisms respectively (Fig. 4). The control architecture is implemented using the RT-Lab based on MATLAB/Simulink running under WinXP. The master and slave are under impedance control in these tests but many other combinations can be made for different tests [2]. All components of the experiments are interfaced by a computer with real-time operating systems (QNX and VxWorks) at a sampling rate of 1kHz. The connection between

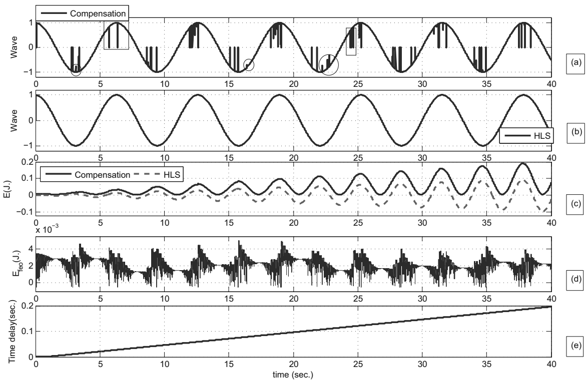


Figure 3: Simulation of communication channel and comparison of energy. (a) Compensated Wave. (b) Wave HLSing. (c) Energy of the channel using HLS and Compensation. (d) $E_{feo.}$, the forward energy observer. (e) Time delay function.

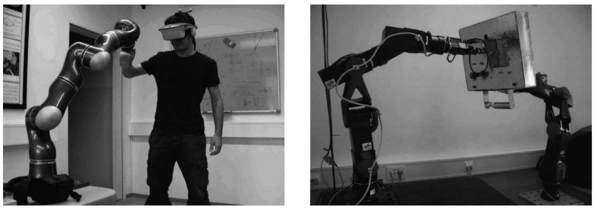


Figure 4: Telepresence test-bed. Operator moves DLR-LWR-III as the haptic display (left side). DLR-LWR-II (slave) performs operation on a movable satellite board (right side).

computers is implemented using UDP sockets. Different time delays and packet drop conditions can be simulated. The control strategy governing the closed loop uses the wave variables control approach so the system is immune to unstability in presence of any arbitrary constant time delays of the communication channel. To cope with the effects of time varying delays and packet-loss our compensation algorithm is implemented at each receiver site. In these experiments the backward delay and losses are ignored for the sake of simplicity.

As it is shown before the proposed compensator essentially does not need a common clock (global time) for master and slave robots which is hard and time consuming goal to achieve specially in long range applications. However for the experiments and for visualisation purposes (to observe the master and slave positions and forces on a synchronised time axis) the Network Time Protocol (NTP) has been used.

In Fig. 5 and Fig. 6 the test results from x-Axis of the master/slave robots are presented. As one can see in Fig. 6(b) a high level of mean time delay (0.7 sec.) and variations (between 0.5 and 0.9 sec.) are selected and the packet-loss rate is also very high (50%). The passivity of the channel is preserved and the operation stability is guaranteed. Some tracking drift is observed in Fig. 5(a) and is due to energy dissipations during the compensator operation (compensations are based on *online forward observed energy* equations and that leads to more energy dissipation than needed). The slave robot experinces a collision with the stiff environment (at ca. 7 sec.) and two longer periods of contact with the stiff environment (satellite board) (during ca. 12-17 sec. and 34-38 sec.). The contact forces of the environment are reflected to the operator's hand.

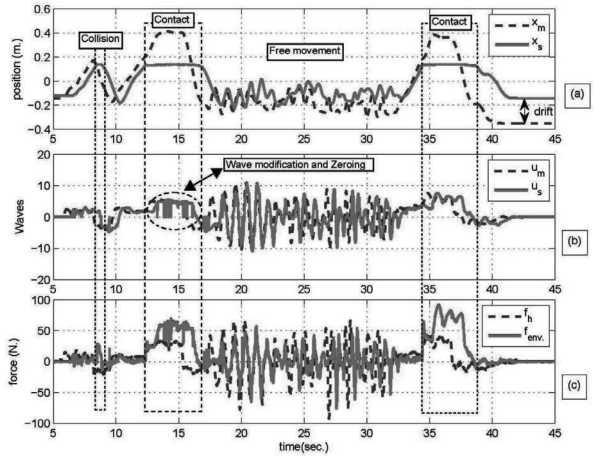


Figure 5: Test with time varying delay and 50% packet-loss in FWD link. (a) Position tracking (x-Axis). (b) Right moving waves (u_m, u_s). (c) Forces (x-Axis).

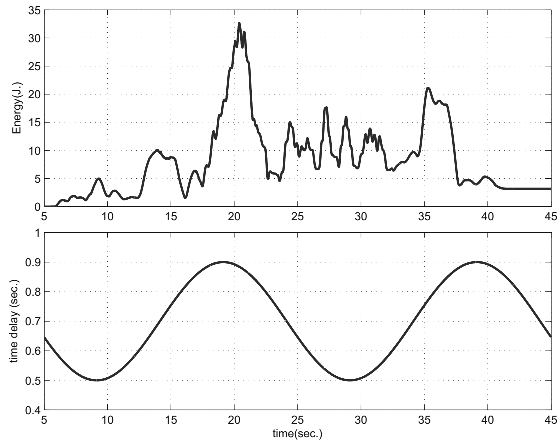


Figure 6: Test with time varying delay and 50% packet-loss in FWD link. (a) Energy balance of the channel. (b) Function of the time varying delay.

6. Conclusions

The compensation method introduced in this paper provides a passivity framework for designing master/slave teleoperation systems over communication networks such as the Internet. The method applies a series of easy to implement modifications to the basic wave variable method. Computer simulations and real tests show the potentials of the method to cope with network non-idealities even in extreme conditions. This energy compensation scheme is based on online virtual energy observation and preemptive activity dissipation. In particular, a passive wave correction under varying time delay circumstances ensures stable and safe operation without being overly conservative using solely Zeroing. Furthermore, holding last sample as primary strategy when a communication packet is lost certainly assures maximum transparency without compromising passivity, since the same algorithm will correct any possibly generated energies in the next sampling time.

Future work will compare the alternative proposed method of the bilateral time domain passivity [3] by quantitatively measuring transparency [15]. The transparency measurement for teleoperation systems with time-varying delays and packet-loss has not been tackled properly yet. Defining perception rates and a general framework to assess transparency of such systems is a potential future direction in this field. Moreover a new approach for reducing the tracking drift of the operation should be sought to make this method a viable and safe way for teleoperation over the Internet and satellite links.

7. Acknowledgments

First author would like to thank Simon Schätzl, Philipp kremer and Thomus Hulin for their help through the project as well as anonymous Enactive reviewers who provided comments on this work.

References

[1] R. Anderson and M. Spong. Bilateral control of teleoperators with time delay. *Automatic Control, IEEE Transactions on*, 34(5):494–501, 1989.

[2] J. Artigas, P. Kremer, C. Preusche, and G. Hirzinger. Testbed for telepresent on-orbit satellite servicing. In *in Proceedings of the Human-Centered Robotic Systems Conference (HCRS)*, Munich, Germany, 2006.

[3] J. Artigas, C. Preusche, G. Hirzinger, G. Borghesan, and C. Melchiorri. Bilateral energy transfer in delayed teleoperation on the time domain. In *in Proc. of the IEEE International Conference on Robotics and Automation (ICRA 2008)*, Pasadena, CA, USA, 2008. IEEE.

[4] P. Berestesky, N. Chopra, and M. Spong. Theory and experiments in bilateral teleoperation over the internet. In *Proceedings of the 2004 IEEE International Conference on Control Applications.*, volume 1, pages 456–463 Vol.1, 2004.

[5] N. Chopra, M. Spong, S. Hirche, and M. Buss. Bilateral teleoperation over the internet: the time varying delay problem. In *Proceedings of the American Control Conference*, volume 1, pages 155–160, 2003.

[6] K. Hashtrudi-Zaad and S. E. Salcudean. Analysis of control architectures for teleoperation systems with impedance/admittance master and slave manipulators. *The International Journal of Robotics Research*, 20(6):419–445, 2001.

[7] S. Hirche and M. Buss. Packet loss effect in passive telepresence systems. In *43rd IEEE Conference on Decision and Control*, pages 4010–4015, December 2004.

[8] P. F. Hokayem and M. W. Spong. Bilateral teleoperation: An historical survey. *Automatica*, 49(12), December 2006.

[9] A. Hosseini. Bilateral telerobotics in presence of time-varying delays and packet-loss. Technical Report 515-08-03, Institute of Robotics and Mechatronics, German Aerospace Center (DLR), February 2008.

[10] D. Lawrence. Stability and transparency in bilateral teleoperation. *Robotics and Automation, IEEE Transactions on*, 9(5):624–637, 1993.

[11] R. Lozano, N. Chopra, and M. Spong. Passivation of force reflecting bilateral teleoperators with time varying delay. Mechatronics'02, Entschede, Netherlands, June 24-26 2002.

[12] C. Melchiorri. Robotic telemanipulation systems: An overview on control aspects. *IFAC*, 2003.

[13] S. Munir. *Internet-Based Teleoperation*. PhD thesis, George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, March 2001.

[14] G. Niemeyer. *Using Wave Variables in Time Delayed Force Reflecting Teleoperation*. PhD thesis, Massachusetts Institute of Technology, Department of Aeronautics and Astronautics Engineering, Cambridge, MA, 1996.

[15] M. Radi, J. Artigas, C. Preusche, and G. Hirzinger. Transparency measurement of telepresence systems. In *Eurohaptics2008*, Madrid, Spain, 2008.

[16] Y. Yokokohji, T. Imaida, and T. Yoshikawa. Bilateral teleoperation under time-varying communication delay. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS '99.*, volume 3, pages 1854–1859 vol.3, 1999.

Development of a robotic platform for cognitive studies on human navigation based on optical flow

Umberto Olcese[1*], Cornelia Beck[2*], Alberto Montagner[1], Stefan Ringbauer[2], Rita Almeida[3], Antonio Frisoli[1], Heiko Neumann[2] and Massimo Bergamasco[1]

[1] *PERCRO Laboratory, Scuola Superiore Sant'Anna*
Piazza Martiri della Liberta' 33 - 56127 Pisa, Italy
[2] *Institute for Neural Information processing, Ulm University*
Oberer Eselberg - D-89069 Ulm, Germany
[3] *Computational Neuroscience Group, Universitat Pompeu Fabra*
Passeig de Circumval·lació 8 - E-08003 Barcelona, Spain

E-mail: a.frisoli@sssup.it, heiko.neumann@uni-ulm.de

Abstract

The development of a robotic platform for cognitive studies on human navigation based on optical flow is presented in this paper. The platform can interact with both a virtual and a real environment, by calculating the optical flow and detecting targets and obstacles. This information is employed by humans to navigate in the environment, although the control laws and their hierarchical composition underlying such a behavior are not known. The proposed platform will provide a powerful tool for addressing these issues.

1 Introduction

Optical flow is an important source of visual information that arises on the retina of a moving person [12]. While in motion, the observer has the task of perceiving his or her immediate environment and to optimally combine this information to accurately guide on-going and future behaviour. The human visual cortex comprises a highly complex hierarchy of visual sensory areas that include the primary visual cortex that serves as input stage that provides input to higher order areas devoted to more complex processing. This processing cascade segregates into two major processing streams [7, 8]. One of these streams is referred to as the dorsal stream and its primary concern is the mapping of objects in space and the analysis of the position and displacement of these objects relative to each other and to the observer. Dorsal-stream areas contribute to motion perception and low-level cognitive processes related to object segmentation and attentional tracking. The other stream, referred as ventral stream, is directed to the inferotemporal cortex and is involved in object recognition. Fast feedforward processing within and across both streams can support object recognition in moving scenes.

Our research activities represent a combined experimental and modelling approach that aims at revealing the mechanisms of visual cognition as they interact with processes of neural decision-making and attention selection. The project aims at integrating cognitive processes, such as decision-making, the relation between decision-making and learning, attention selection based on such decisions, and the coordination of sensory processes with selection processes in the moving participant. Such issues can be explored by creating a sophisticated robotic platform with independent sensory and motor abilities to test neural-network models, such as, e.g., visual search and obstacle avoidance in real scenes. This paper focuses on the actual implementation of such a robotic device, both from a hardware and software point of view. In particular, during the robot software architecture creation, the general scheme shown in Figure 1 has been followed and replicated.

A robotic platform has been developed in order to form a test bed for the investigation of perception and

*These authors contributed equally to the manuscript.

Figure 1: Schematic functional architecture. The modules we implemented focused on different steps of the control loop. The optical flow calculation module worked at the V1/V2-MT/MST level, the SpikeNet object detection module at the V4-IT level and the decision module at the PFC-PMC/MC level.

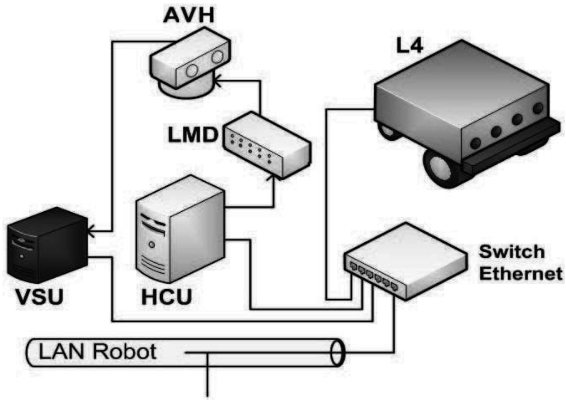


Figure 2: Integrated system general architecture

action interaction. The robot itself provides the motor-component, whereas the sensory part of the robotic system consists of two cameras. The software implementation of the visual system is derived from the human cortex structure, i.e. replicating the aforementioned division into dorsal and ventral pathways. Mainly the information of the optic flow is analysed but also information of the form channel is used via cross interaction.

A neural architecture is developed with realistic components of the primary visual pathway and the combination with a simple motor system. Hereby, a decision module (prefrontal cortex) is integrated, that influences the behaviour. Altogether, we want to simulate and investigate behaviour for tasks of different complexity, where the results generated in the single layers are passed to the decision layer.

2 Platform architecture

Figure 2 outlines the overall architecture of the mobile platform. In particular, a custom Active Vision Head (AVH) has been installed onto a commercial mobile platform. The mobile device has two active wheels and is a differentially-driven robot, i.e., differences in wheel velocities lead to robot self-rotations. The AVH is capable of simulating typical human eye movements, such as fast ballistic, saccadic and slow pursuit movements, with comparable performance in terms of response time, smoothness of motion and trajectory tracking, and with independent sensory and motor abilities.

2.1 Hardware details

The chosen mobile platform is the Labo-4, developed by AAI, Canada. Such platform is a high-payload (up to 30kg) device which can be configured for a wide range of experiments and real-world applications. Despite the fact that eight ultrasonic and eight infrared sensors are mounted and integrated on the platform, they are not currently employed for navigation tasks, which are required to be completely vision-based.

The Active Vision Head (see Figure 3) is designed in order to simulate pan and tilt movements of the head itself and the eyes. Head pan and tilt movements, eye tilt and independent eye pan movements can be controlled on-line during while operating the robot. The presence of both eye independent pan movements can be exploited in order to control the eye vergence in a symmetric or asymmetric way. Every Degree of Freedom is actuated by means of DC motors either directly connected to the joint or connected to specifically designed transmission systems.

2.2 Software details

In order to comply with the testing of new computational models, requiring the fulfillment of execution in a given computational time to generate an appropriate optical flow in the observed scene, the robotic platform and the vision head are controlled directly by the main control unit. The robot navigates in a real scenario, and the optical flow is generated by the effective movement of the robot. This requires that the time performance for the execution of the control loop should be fast enough to allow the decision making of the robot. The proposed architecture for the robot tries to comply with the aforementioned division of the visual processing in humans into dorsal and ventral pathways, by implementing a double-stream parallel processing with

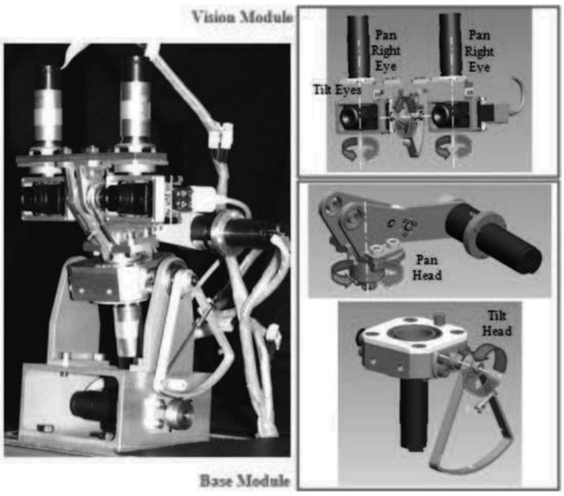


Figure 3: Active Vision Head

very specific functionalities:

- multithreading support: deployment of the application over multiple threads to guarantee parallelization of processes;
- deployment of code execution on a quad-core processor machine for the optimization of speed of execution of numerical code;
- adoption of dll modules for the integration of modules and code sharing.

3 Control architecture

In accordance to the aforementioned general guidelines, two concurrent processes to perform in parallel optical flow computation and target detection have been implemented (see Figure 4). On the one hand, optical flow computation is mainly used for a corridor centering task, i.e., a navigation task in which the robot is required to keep in the middle of a straight or curved corridor. On the other hand, predefined objects can be detected in the scene through a SpikeNet module, a commercial software module (www.spikenet-technology.com) for neuroinspired fast object detection. Such module can be used either for targeting or avoiding the detected object.

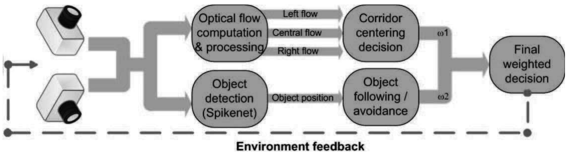


Figure 4: Control system general architecture

3.1 Optical flow computation

The optic flow used for the navigation is computed by a model from Bayerl&Neumann [1] that is based on the neural processing in our visual system. This model simulates the first stages of motion processing in the human dorsal pathway, namely V1 and MT. In model area V1 a first detection of optic flow is realized, model area MT estimates the optic flow of larger regions. Both model areas contain feedforward, feedback and lateral connections that form a three-level processing cascade. The key components of this model are the interplay of the two different areas in two different spatial scales that are recurrently connected. In addition, the lateral connections lead to a normalization via shunting competition that strengthens the activity of unambiguous motion signals.
As the navigation in real-time requires very fast and stable optic flow computation, the model is calculated in an algorithmic way [2] that is accelerated due to a sparse representation of neurons and a very efficient initial motion detection stage using an extended version of the Census Transform [11]. Optic flow stabilization is, amongst other, achieved using to temporal integration and a compensation of the rotational components that arise due to jittering of the moving robot.

3.2 Control algorithms

The robot was at first equipped with a basic control strategy to provide a basic level of visuo-motor behavior. The robot's basic task is to safely navigate through passages of empty space between potentially hazardous obstacles. Consider, for example, a corridor that is created by two walls of approximately parallel orientations which remain stationary while the robot (observer) moves. Since the projected 3D visual flow vectors leading to 2D image flow are inversely scaled as a function of the scenic depth structure, a balancing strategy can be easily derived using peripheral flow information. Through balancing the average motion in both peripheries guarantees steering the robot along a path approximately along the middle of a corridor-like environment. Basic optical flow based control strategies try to balance left and right flow values, in order to keep the robot in the middle of a corridor-like environment [3, 6, 9, 13, 10]. The approach that was employed in this work was to use the average flow value in two areas in the left and right regions of the field of view as inputs to the control system. A third area, located in the middle of the field of view, was employed for calculating the mean rotational component, to be subtracted from all other flow values. If the left flow is, for example, greater than the right one, the robot is closer to the left wall and will have to move to the right in order to keep in the center. Since depth structure is encoded directly in the translatory flow component, the control mechanism has an implicit sense of relative depth which finally needs to be balanced. From this basic idea all the following control algorithms were developed.

3.2.1 Standard control system

A finite state controller was developed as a first means to control the robot. Two values are computed from the left and right optical flows [6]: the difference d, corresponding to the difference of the instant absolute values of left and right optical flows; a threshold value th, corresponding to the sum of the instant absolute values of the left and right optical flows divided by a fixed normalization parameter k.

The d value is directly correlated to the distance between the robot and the corridor centerline to be followed. Positive values correspond to the robot being closer to the left wall and vice versa. The greater $|d|$ is with respect to th and the closer the robot will be to a wall. Two thresholds were employed for choosing between a small or a large robot rotation, necessary to move toward the center of the corridor. After each rotation command no output is given to the robot for some time steps (usually 2), in order to let the flow stabilize to reliable values. See Figure 5 for a block diagram of the proposed control architecture.

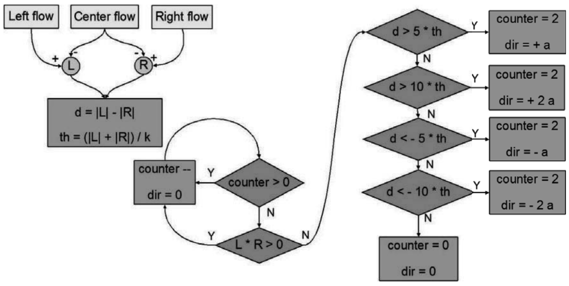


Figure 5: State-machine control algorithm

The sequence of output command has therefore a spike-like appearance, each rotation being separate at least two steps from the subsequent one. A control strategy employing continuous rotations lead to a more oscillating behavior.

3.2.2 Two-pool neural control system

A neuro-inspired decision module [4, 5] has then been integrated on the robotic platform. This module consists of two self inhibiting neural pools (figure 6). Inputs to the system are the left and right flows. The output command was the difference between the activities of the two pools. Compensation of self rotation and additional variables for allowing flows to stabilize after every rotation command were still employed. The equations which regulate the dynamic behavior of the two neural pools are as follows:

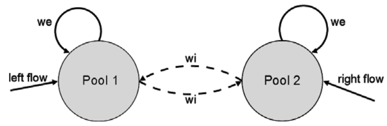


Figure 6: Two-pool neural control architecture scheme

$$\begin{cases} \dot{\omega_1} = -\omega_1 + \frac{a}{1+e^{-b(w_{11}*\omega_1 - w_{12}*\omega_2 + \lambda_1 - c)}} \\ \dot{\omega_2} = -\omega_2 + \frac{a}{1+e^{-b(w_{22}*\omega_2 - w_{21}*\omega_1 + \lambda_2 - c)}} \end{cases} \quad (1)$$

where ω_1 and ω_2 are the two neural populations, λ_1 and λ_2 are the two inputs (flows), w_{11} and w_{22} are self-excitatory weights, w_{12} and w_{21} are reciprocally inhibiting weights, a, b and c are system parameters.

The potential field of such a system is designed in such a way that, when the two inputs are balanced, there are two symmetrical potential wells, corresponding to left and right rotation. Thanks to random oscillations in the activity of the network the state of the system can fluctuate between the two wells and, more often, stabilize in an additional equilibrium point between the two wells, leading therefore to no rotation. When one input is greater than the other, one potential well becomes deeper than the other and the system is strongly biased toward a rotation. The parameters of this system have been slightly modified in order to obtain a better robot behavior.

3.2.3 Three-pool neural control system

The two-pool control system represents a first step toward a fully biological navigating device. The fact setting the output to zero for some steps after each rotation command is still present and represents an artificial component.

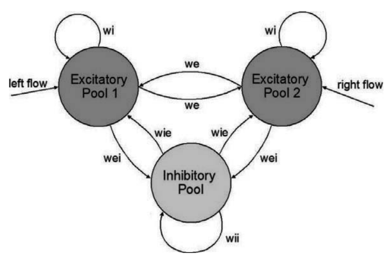


Figure 7: Three-pool neural control architecture scheme

By adding a third inhibitory neural pool (see Figure 7) it was possible to obtain a fully functional neuroinspired control system. The main function of this pool would be to inhibit the other two after each rotation command, so that flows could autonomously stabilize.

3.3 Object recognition

SpikeNet is an image recognition and tracking technology using networks of asynchronous spiking neurones. SpikeNet uses processing algorithms that are directly inspired by the strategies used by the human visual system which outperforms even the most sophisticated machine vision systems. Indeed, the human visual system is able to analyse a complex scene in a fraction of a second.

In order to be used for object detection, the module needs to be trained by providing one or more images of the object to be detected in the scene and appropriately tuning threshold parameters. With a proper parameter tuning, it is possible to reduce computational burden and recognize objects even in difficult conditions. The SpikeNet routine has successfully been integrated with the robotic platform software in order to perform parallel object detection and navigation tasks using the multithreaded software architecture. In particular, additional control algorithms for object tracking and object avoidance have been introduced in the software architecture.

Object tracking was achieved by adding a rotation command to the one calculated with the optic flow. No output is given if the object is in the central part of the view field, in order to avoid oscillations. It is also possible to detect the distance of the object by estimating its size. The robot was able to move without hitting side walls and, at the same time, perform object tracking.

In order to obtain obstacle avoidance a control law complementary to the one employed for object tracking was designed. If the obstacle was detected in the outermost left and right part of the flow field, no rotation was necessary. Corridor centering and concurrent obstacle avoidance were successfully achieved.

4 Preliminary results

With a typical input sequence containing left and right peripheral flows (e.g. Figure 8, which corresponds to a typical acquisition sequence in a real navigation task) the typical output of the proposed control architectures is a spike-like sequence (see Figure 9 for a sample result comparing the three different control strategies outputs for a navigation task on a same corridor). Such control signal has proven its effectiveness in keeping the robot in the center of the corridor, without having an oscillatory behaviour. As a matter of fact, the robot was

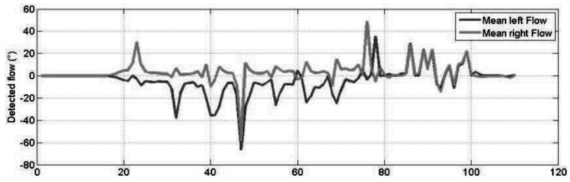


Figure 8: Sample input sequence

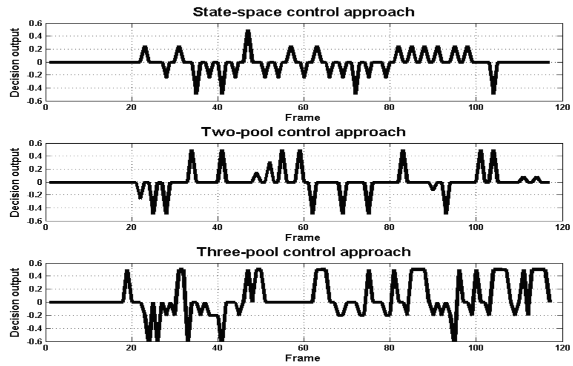


Figure 9: Sample output sequences

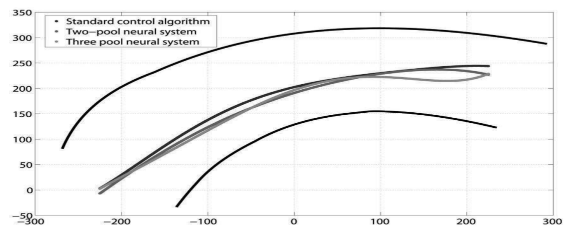


Figure 10: Average trajectores recorded using the different control systems

able to move in a straight or curved corridor without hitting side walls and able to correct the trajectory in order to avoid the collision with walls (see Fig. 10 for the average trajectories recorded employing the three control strategies). Avoidance of large objects - such as a person - was also possible by the detection of the change in optical flow, without changing any control parameter. The presence of a large obstacle in one side of the flow field causes indeed an asymmetry between left and right mean values, hence resulting in a steering command. Moreover, concerning the usage of the two-pool neural algorithm, it has been shown that, after accurate parameter tuning, the robot is able to move in both highly and low textured corridors. Obstacle avoidance based on optical flow only is also possible for extended objects. The robot is able to pass a corridor like environment with sidewise obstacles forming a narrow zig-zag track. In such a setting the rotation compensation mechanism has proved to be indispensable.

5 Discussion

This rather basic centering behavior to control the robot's route can be considered as an online decision-making process in which asymmetries in the peripheral flow distributions above threshold lead to corrections in the route-selection process. This is continuously monitored by comparison between flow information generated along the model dorsal pathway of motion computation.

The robot hardware and software configurations have been tested in a navigation task where the robot had to move in a corridor-like environment, avoiding walls and obstacles and reaching a target. Both the neuromorphic module for optical flow calculation and the Spikenet routines for object recognition have proved to be effective tools for performing visual based navigation. Several control strategies have been implemented, with the aim of developing a fully neuroinspired control architecture.

It has been shown that human navigation is influenced by the presence of an optical flow component [12], but few data are available on the control laws which our nervous system employ. As demonstrated a technical base driven by cognitive neuronal mechanisms opens the ability to reproduce and evaluate such empirical findings. Therefore our platform, designed with a bioinspired approach, represents a valuable tool for performing experiments meant to investigate the existing neuroscientific hypotheses.

6 Acknowledgments

This work is supported by EU FP6 IST Cognitive Systems Integrated project DECISIONS IN MOTION. The authors would like to thank Simon Thorpe and Edoardo Sotgiu for their valuable collaboration.

References

[1] P. Bayerl and H. Neumann. Disambiguating visual motion through contextual feedback modulation. *Neural Comput*, 16(10):2041–2066, Oct 2004.

[2] P. Bayerl and H. Neumann. A fast biologically inspired algorithm for recurrent motion estimation. *IEEE Trans Pattern Anal Mach Intell*, 29(2):246–260, Feb 2007.

[3] D. Coombs, M. Herman, T. Hong, and M. Nashman. Real-Time Obstacle Avoidance Using Central Flow Divergence, and Peripheral Flow. *IEEE Transactions on Robotics and Automation*, 14(1):49, 1998.

[4] G. Deco and D. Martí. Deterministic analysis of stochastic bifurcations in multi-stable neurodynamical systems. *Biological Cybernetics*, 96(5):487–496, 2007.

[5] G. Deco, L. Scarano, and S. Soto-Faraco. Weber's Law in Decision Making: Integrating Behavioral Data in Humans with a Neurophysiological Model. *Journal of Neuroscience*, 27(42):11192, 2007.

[6] A. Duchon and W. Warren. Robot navigation from a Gibsonian viewpoint. *Systems, Man, and Cybernetics, 1994.'Humans, Information and Technology'., 1994 IEEE International Conference on*, 3, 1994.

[7] D. J. Felleman and D. C. V. Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex*, 1(1):1–47, 1991.

[8] J. V. Haxby, C. L. Grady, B. Horwitz, L. G. Ungerleider, M. Mishkin, R. E. Carson, P. Herscovitch, M. B. Schapiro, and S. I. Rapoport. Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proc Natl Acad Sci U S A*, 88(5):1621–1625, Mar 1991.

[9] J. Santos Victor, G. Sandini, F. Curotto, and S. Garibaldi. Divergent stereo for robot navigation: learning from bees. *Computer Vision and Pattern Recognition, 1993. Proceedings CVPR'93., 1993 IEEE Computer Society Conference on*, pages 434–439, 1993.

[10] J. Serres, F. Ruffier, S. Viollet, and N. Franceschini. Toward optic flow regulation for wall-following and centering behaviors. *International Journal of Advanced Robotics*, 3(2):147–154, 2006.

[11] F. Stein. Efficient computation of optical flow using the census transform. *DAGM Symposium 2004*, pages 79–86, 2004.

[12] W. H. Warren, B. A. Kay, W. D. Zosh, A. P. Duchon, and S. Sahuc. Optic flow is used to control human walking. *Nat Neurosci*, 4(2):213–216, Feb 2001.

[13] K. Weber, S. Venkatesh, and M. Srinivasan. Insect inspired behaviours for the autonomous control of mobilerobots. *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, 1, 1996.

Mechanical Design of Novel Wearable Two-Finger Haptic Interface System

Marco Fontana, Federico Tarri, Simone Marcheschi, Fabio Salsedo, Massimo Bergamasco
PERCRO Laboratory, Sant'Anna School of Advanced Studies, Pisa, Italy
E-mail:[fontana, federico.tarri, marcheschi, fabio, bergamasco]@sssup.it

Abstract

This paper describes the design process and the technical development of a novel exoskeletric haptic interface system. The realized device is a wearable hand exoskeleton for the index and the thumb fingers and it is equipped with a special designed mechanical tracker for absolute position tracking.

The development of this device finds a place in the framework of the new research branch called fingertip haptic, which consists of investigating on novel devices and principles able to include some tactile features in the traditional kinesthetic devices.

In the first part of the paper we describe the haptic device architecture and performances, showing the design choices that have been made for optimizing accuracy and resolution. In the second section we describe the design process of a mechanical tracker able to measure the position of the hand and to compensate the weight of the haptic device mounted at the end tip.

1. Introduction

Haptic Interfaces (HI) are devices able to establish a bilateral communication between humans and computer through the sense of touch. Despite of their high potential, HI has not broken into the market and their spreading is still limited.

There are several reasons that explain the slow growth of HI market. Basically we can say that there are some concurrent causes like low versatility, insufficient exploitation of the human haptic channel, bulkiness of devices and costs, just to mention some of them. In answer to some of these issues a new research branch called Fingertip Haptics [4] was recently born. The main aim is to conceive new concepts and devices for a richer stimulation of the sense of touch in order to provide devices that are able to simulate the direct contact with the human fingers.

Within this research branch many devices have been developed in different laboratories. Yokokohji from Japan is carrying on a research line started by Hirota and Hirose [5], [6], working on a new class of haptic devices called Encountered Type Haptic Interfaces [7]. This novel kind of haptic interface is able to reproduce the non-contact-to-contact transition; in particular the HI doesn't keep the contact with the user hands during free space movements but just follows the limbs with a non-contact tracking system and the contact takes place only when the virtual surface is touched.

Hannaford from Washington University is working with an experimental setup and he is conducting some early experiments on "small scale forces haptic feedback" [8]. His objective is to put the bases for the realization of a low power consuming haptic device.

In this paper it is described the design work that has been carried out for building an Exoskeletric Haptic Interface (EHI) for the hand. An EHI is a haptic device whose mechanical structure is located in proximity of the part of the body that is interested in the interaction. The kinematics of such devices is often designed in a way that the linkages of the structure follow the movements of the human operator. EHIs have attracted a lot of attention among the haptic research community since they show several advantages. Their most relevant feature is the optimal workspace matching between the human reachable workspace and the haptic device workspace. Moreover, EHIs perfectly fit the applications where reduced visual and spatial encumbrance is needed.

This paper describes a system that has been realized addressing the issue of direct contact interaction. The whole system is composed by a wearable Hand Exoskeleton (HE) able to exert forces on the user index finger and thumb fingertips. This device is mounted on the dorsal side of the palm of the user right hand. The HE is coupled with a mechanical tracker whose function is to detect in real-time the position and orientation of the hand and to compensate the weight of the haptic device through a counterbalancing system.

The main aim of the design work that has been carried out was to obtain a high performance device able to exert force in the range of a few Newtons but with exceptional capabilities of accuracy and resolution. The target of application of the HE that is described in this paper is the haptic interaction and manipulation through the fingertips of the index and thumb. The reference tasks are the precision grasping between index and thumb, manipulation and exploration of surfaces and objects involving forces in the range of 0-5N.

2. The haptic interface

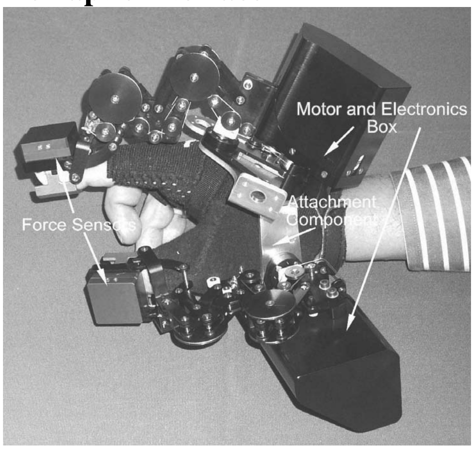


Fig. 1: The Hand Exoskelton (HE)

The Haptic device that has been developed in this work is represented in Fig.1. It is a HE for the index and thumb finger. The device is wearable and can be mounted on the hand of the user through a glove-like interface and it is able to exert a desired force on the fingertips of the index and thumb of the user along any direction.

The choice of a wearable anthropomorphic solution was determined essentially by the advantages that such architecture shows when multi-finger haptic feedback is requested. These advantages include low mass of the mechanics, lack of limitation of position and orientation workspace, intrinsic avoidance of internal collision between the different arms of the haptic device, etc...

2.1 Kinematics

The kinematic scheme that has been adopted for the developed HE can be defined as a Quasi-Anthropomorphic Kinematics. The choice was driven by the fact that anthropomorphic kinematics presents many positive features that well fit the requirements of a high accuracy and high resolution haptic device:

- Highest ratio between workspace of the mechanism and required workspace;
- Short path for the robotic linkages;
- Wearability with limited visual and spatial encumbrance.

A quasi-anthropomorphic kinematics means that the kinematics is morphologically identical to the finger kinematics but it slightly differs for the length of the links. The perfect coincidence would be desirable but it is not allowed because when the flexion joints are extended the finger kinematics falls in a singular position of the manipulator.

The concept of quasi-anthropomorphic kinematics is represented in Fig.2 where the model of the finger and the quasi-anthropomorphic haptic device are shown together.

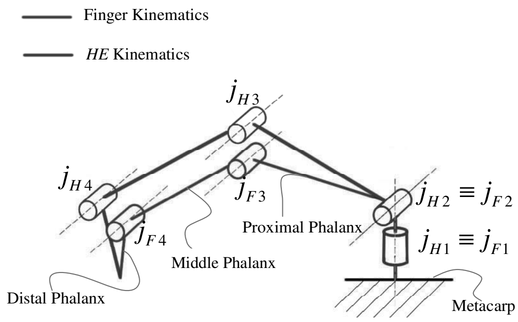


Fig. 2: Scheme of the kinematics

In order to reduce the number of actuators we assumed a simplification in the kinematic scheme. The kinematics of the haptic device represented in Fig.2 has 4 DoFs (as many as the finger), but we are interested in generating a 3 DoF force on each fingertip. The actuation of 4 DoFs would require the use of four actuators. Since the actuators usually constitute the biggest fraction of the weight of the mechanical assembly of an HI, we chose to introduce a simplification that allows the use of only three motors. With reference to Fig.2, the rotation between the distal and middle phalanx (joint j_{H4}) has been coupled with the rotation between the middle and the proximal phalanx (joint j_{H3}). This coupling reflects the behavior of the human finger and in particular it has been assumed that the rotation of the joint j_{F3} is equal to the rotation of the joint j_{F4}. The implementation of such feature has been realized through a custom designed mechanism that is described in the following section.

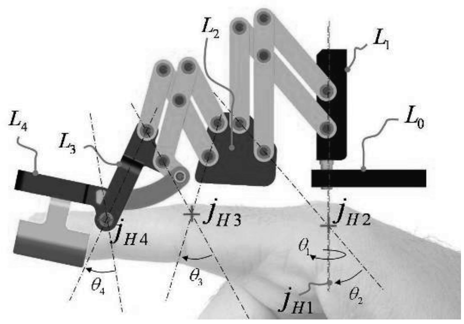


Fig. 3: Scheme of the whole kinematics for the index finger

2.2 Mechanical Implementation

The scheme of the implementation of the kinematics is represented in Fig.3.

j_{H1}: ab/ad-duction joint

The first joint (j_{H1} in Fig.3) of the kinematic chain is implemented in a simple way, with a pin supported by two rotational bearings. The axis joint is oriented along the normal to the plane of the hand palm. The avoidance of collision with the user finger is achieved locating the pin and the bearings on the dorsal side of the hand.

j_{H2} and j_{H3}: flexion-extension joint

The flexion-extension joints j_{H2} and j_{H3} have to be carefully designed, since the collision with the finger may occur in many different ways. A possible solution is to place the joints j_{Hi} at the side of the finger like in [1]. In this solution, the joints with their bearings are located on both the sides of the finger. This implementation is functional and reliable but shows evident inconvenience due to the lateral encumbrance of the joints.

The implementation that we adopt makes use of *Remote Center of Motion* (RCM) systems. RCM are mechanisms that are able to implement the rotation of a body around a fixed axis that is remotely located from the structure of the joint. There are several mechanisms able to implement this feature. Many of them are reported and analyzed by Pei [2] and Fiaschi [3]. Fiaschi has analyzed several RCM for the implementation of flexion joints of hand exoskeletons. The analysis consists of a score ranking that considers the complexity of implementation, the encumbrance and the mechanical characteristics.

The simpler and more efficient RCM is composed by two parallelograms connected together to form a 6-bar mechanism as in Fig.4. Such mechanism allows the *Link i* to rotate around the Remote Center of Rotation (RCR) of the angle θ_i .The mechanism has a singularity-free stroke within the limits:

$$0 < \theta_i < \pi$$

that is acceptable since statistically the stroke of the rotational joints of the human hand does not exceeds that values. If a RCM is used, instead of lateral bearing joints, it is possible to locate all the mechanism of the HE on the dorsal side of the finger (obviously except for the last link). This mechanism is chosen for the implementation of the joints j_{H2} and j_{H3}.

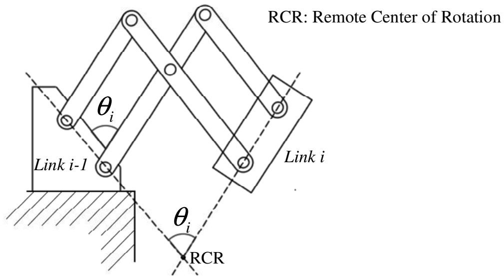


Fig. 4: Scheme of the remote center of the RCM

Joint j_{H4}

The j_{H4} could be implemented with the same RCM joint. However the choice was to implement a joint with lateral bearings. The choice was commanded by the need of leaving the dorsal side of the last link free to possibly host a tactile actuator.

As mentioned in the previous section the joint j_{H4} has been coupled with the joint j_{H3}. This has been done for reducing the number of actuators and consequently the weight of the device to the minimum needed for an acceptable use. This approximation is largely acceptable for the index, middle, ring and little finger. This is not valid for the thumb, but in order to simplify the design and allow reaching an acceptable weight of the whole system it was decided to have a coupling of the proximal-interphalangeal (PIP) and distal-interphalangeal (DIP) joints also for the thumb.

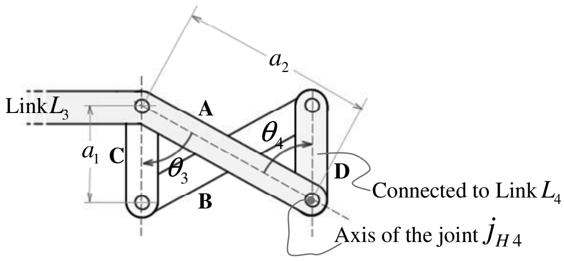


Fig. 5: Crossed parallelogram mechanism

The coupling between the joints j_{H4} and j_{H3} has been implemented avoiding the use of toothed wheels and gears in order to reduce the backlash and friction at best. A crossed parallelogram mechanism like the one depicted in Fig.5 was used for such purpose.

Transmission and Actuation

In serial kinematics robotic device the actuators are rarely located directly on the axis of the joints because their own weight would bring to waste the available motor torque for gravity compensation. Also in our design the actuators are delocalized and the joints are actuated through a cable-pulley transmission system that conveys the torques to the joints of the device.

The three actuators are located on the first moving link of the structure. The choice was a trade-off between complexity of the transmission and the best solution for the gravity compensation.

A patent application has been filed for the transmission system.

2.3 Force sensing and control electronics

In order to enhance the force reflecting accuracy of the HE, a custom designed force sensor has been integrated at the end-effector of the device. The force sensor is a strain gauge based device that makes use of the known Maltese-Cross architecture. It is able to measure a force applied on the fingertip of the user (see Fig.1) within a range of 0-5N with a resolution of about 14bits, given by the maximum noise.

In order to minimize the electrical noise and to simplify the cabling of the device, the HE was equipped with a custom on board electronics for the conditioning of force signal and acquisition.

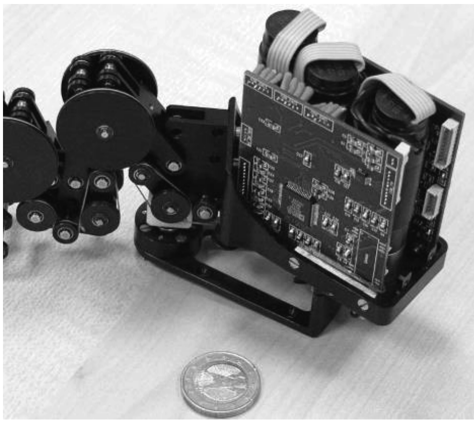


Fig.6: Picture of the electronics integrated in the mechanics

The whole electronics components were integrated into the mechanical design of the HE (see Fig.6). The noise that could affect the analog signal is than minimized since the length of the cable that carry analog signal is reduced. Moreover a simplification of the cabling is achieved since the whole data are exchanged through a couple of T_x-R_x cables and only two power cables are routed for the power supply.

2.4 Performances

We report the mechanical performances of the HE in brief, verified sperimentally:

- Maximum Force: 5N in worst condition over any abitrary direction;
- Force Resolution: 0.01N (depends on the controller);
- Position Resolution: 0.05mm;
- Weight of the assembly: 1.1kg;
- Mechanical Bandwidth: 23Hz.

3. Mechanical Tracker

In virtual reality applications the forces that have to be transmitted to the user are calculated according to the absolute position of the haptic interface end-effectors.

A portable interface attached to the user hand gives only the relative position of the fingertips respect to the hand. A tracking system that measures the position and orientation of the user hand is then needed in order to make the device usable.

Many position and orientation tracking techniques are available, but since our application involves haptic interaction we requested high resolution and high refresh rate.

After a preliminary analysis of the different tracking technologies, the attention has been focused on the mechanical trackers, but the few commercial mechanical trackers available on the market did not fulfill our requirements.

A custom designed mechanical tracker has been realized, based on a novel magnetic sensing principle. The realized device is basically a 6 DoF passive serial kinematic chain with sensorized joints. The kinematic structure is obtained through a modular design by the serial repetition of a 2 DoF sensing module. The angular sensors are based on a novel Magnetic-Hall Effect sensing principle.

Moreover the developed tracker is able to relieve the weight of the HE, by means of a passive counterbalancing system.

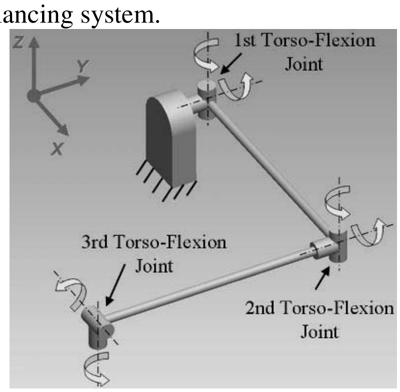


Fig. 7: Scheme of the kinematics

3.1 Kinematics

The kinematic structure of the tracker is represented in Fig 7. The serial kinematic chain is obtained using three units of a 2 DoF joint module that is called Torso-Flexion joint. The Torso-Flexion joint is composed by two rotational joints with orthogonal axes. The mechanical implementation of such kinematics is shown in Fig. 8.

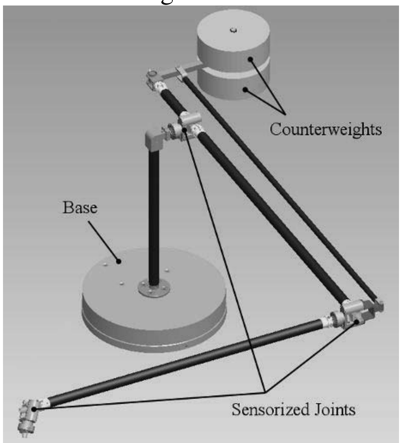


Fig.8: Mechanical implementation of the tracker

The three Torso-Flexion joints are connected through aluminium and carbon fiber tubes, which completely hide the electrical cabling. The workspace of such device is perfectly scalable by substituting the tubes with others of different length.

3.2 Sensing principle

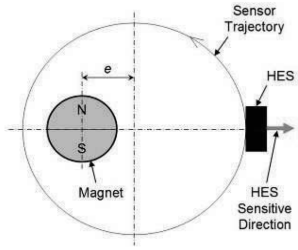


Fig.9: Sensing principle scheme

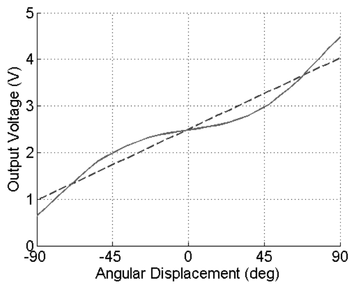


Fig.10: Plot of output voltage from HES (solid line) and of LSL (dashed line) in an experimental test (linearity error is amplified by 5). HES supplied with 5V by a high precision voltage supplier and output voltage measured by a high precision multimeter

The innovative aspect of the proposed tracker is the sensing system. The novel idea is shown in Fig. 9. An hall-effect sensor (HES) follows a circular trajectory around a cylindrical (or annular) magnet magnetized along a diameter. The circular trajectory is not centered on the axis of the magnet, but there is an eccentricity *e*. By tuning this eccentricity it is possible to achieve a good linearity field between the measured component of the magnetic flux density **B** and the angular displacement. In Fig.10 it is shown the good linearity according to the results of an experimental test. The output voltage from the HES respect to the angular displacement (±90º) and the least square line (LSL) are plotted, with the linearity error (difference between real data and LSL) amplified by 5 to distinguish the two curves. The maximum linearity error on the angular range ±90º is 89mV.

The mechanical implementation of the novel sensing principle is shown in Fig.11. A simple cylindrical magnet has been chosen for the flexion joint, while an annular magnet has been chosen for the torsional joint, in order to allow the cables to pass through.

The signals from the HES are acquired by an embedded A/D conversion electronics with 16bits resolution located inside the base of the tracker.

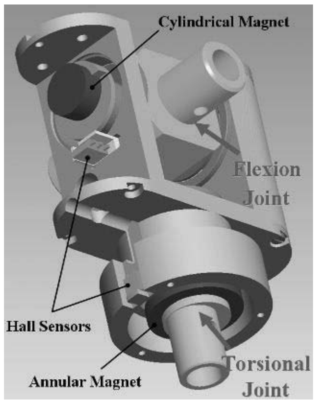


Fig.11: Mechanical implementation Torso-Flexion Joint

3.3 Counterbalancing system

Although the weight of the HE (around 1kg) is sustainable by the user, for application that require long working time it is highly desirable to completely compensate the weight of the device. The developed mechanical tracker has been designed in order to assume this further functionality.

The counterbalancing system is based on a pantograph mechanism with a mass attached to the opposite side respect to the HE (see Fig.8 and Fig. 12). The Fig.12 shows a top view of the whole system; in the picture x-y is an horizontal plane and gravity vector is oriented along the z-axis. The counterbalancing mass is attached at the point E, the HE is attached at the point A. The two triangles ABC and CDE are similar when

$$L_1/L_2 = L_4/L_3$$

Under these hypotheses the two segments AC and CE are aligned during the tracker movements thanks to the parallelogram BDGF. If we indicate with M_1 the mass of HE and M_2 the mass of the total counterweight, the weight of mass M_1 can be balanced with a mass M_2:

$$M_2 = M_1 \frac{L_5}{L_6} = M_1 \frac{L_2}{L_3}$$

As a consequence, if $L_3 \leq L_2$ it is possible to balance the mass M_1 with a mass $M_2 \geq M_1$. For our system we choose $L_2 = L_1/4$, resulting a counterbalance weight of M_2=4 kg.

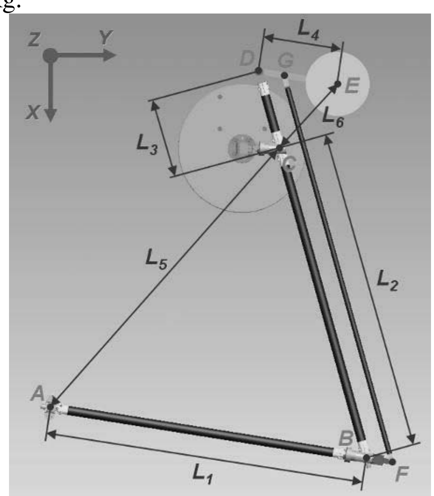


Fig.12: Counterbalancing system principle

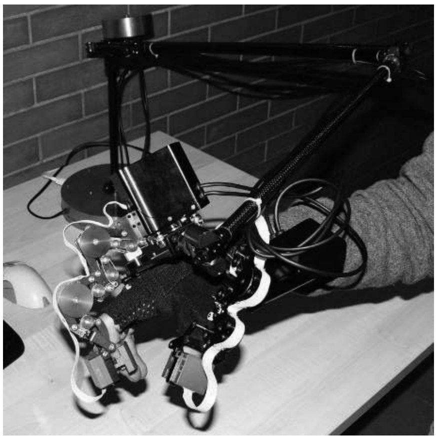


Fig.13: The whole system

4. Conclusions and future works

The HE and tracker has been integrated together as represented in Fig.13. The result was a high performances system able to reach high force accuracy within a large workspace. The system is able to exert forces on the fingertip of the hand allowing the generation of forces that are typical of tasks like precision grasping and surface exploration.

The force range is sufficient for rendering believable interaction forces with rigid objects.

A first preliminary qualitative observation is that the counterbalancing system really improves the quality of the interaction.

Further improvements are foreseen for the attachment system on the user hand. This component is particularly critical since a bad force transmission can destroy the illusion of perceiving external forces on the fingertips.

References

[1] C. A. Avizzano, F. Barbagli, A. Frisoli, M. Bergamasco. The Hand Force Feedback: Analysis and Control of a Haptic Device for the Human Hand, *Proceedings of IEEE SMC 2000, Intern. Conf. on Systems, Man and Cybernetics*, Nashville, TN, USA

[2] X. Pei, J. Yu, S. Bi, G. Zong. Enumeration and Type Synthesis of One-DOF Remote-Center-of-Motion Mechanisms, *12th IFToMM World Congress*, Besançon (France), June18-21, 2007

[3] M. Fiaschi. Analisi e progetto di un'interfaccia aptica di tipo esoscheletrico per la simulazione di prese di potenza e di precisione, *Master Thesis in Mechanical Engineering*, Faculty of Engineering, University of Pisa, Pisa, Italy, 2003

[4] M.A. Salada, J.E. Colgate, M.V. Lee, P.M. Vishton. Validating a novel approach to rendering fingertip contact sensations, *HAPTICS '02, Proceedings of the 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 24-25 March 2002, Page(s):217 – 224

[5] K. Hirota, M. Hirose. Development of surface display, *Virtual Reality Annual International Symposium, 1993 IEEE*, 18-22 Sept. 1993, Page(s):256 – 262

[6] K. Hirota, M. Hirose. Simulation and presentation of curved surface in virtual reality environment through surface display, *Proceedings of the Virtual Reality Annual International Symposium*, 11-15 March 1995, Page(s):211 – 216

[7] Y. Yokokohji, N. Muramori, Y. Sato, T. Kikura, T. Yoshikawa. Design and path planning of an encountered-type haptic display for multiple fingertip contacts based on the observation of human grasping behaviour, *Proceedings of 2004 IEEE International Conference on Robotics and Automation*, Volume 2, Apr 26-May 1, 2004, Page(s):1986 - 1991

[8] R. Leuschke, E.K.T. Kurihara, J. Dosher, B. Hannaford. High Fidelity Multi Finger Haptic Display, *Proceedings of the First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 18-20 March 2005, Page(s):606 – 608

Analysis of Percussion Grip for Physically Based Character Animation

Alexandre Bouënard* † Marcelo M. Wanderley† Sylvie Gibet* ‡

* *VALORIA-Samsara, Université de Bretagne Sud, Vannes, France*
† *IDMIL, McGill University, Montreal, Canada*
‡ *IRISA-Bunraku, Université de Rennes I, Rennes, France*

Abstract

This paper presents the analysis of percussion technique for the simulation of virtual timpani playing situations. The analysis has been processed for two subjects performing with two types of grips (French and German), yielding to the extraction of percussion technique parameters from a motion capture database. These parameters are shown to be necessary for specifying the control and extending the expressivity of the simulation process.

1 Introduction

The present contribution aims at combining the assets of both the dynamic control of virtual characters and the analysis of percussion instrumental gestures, for synthesizing novel physically responsive and sounding instrumental situations. The physical simulation is indeed of great importance for taking into account the physical parameters of the strike (tension, forces) that influence the dynamics of the gesture and the produced sound, while the analysis of percussion grips gives a level of control and different modes of playing to the virtual character.

Either for the physical simulation or the analysis parts, the interest of using motion capture data appears crucial and twofold. Firstly, it provides realistic examples of the human motion under study for driving the virtual character simulation. Physically-based character animation techniques provide a solution to the limited reuse of motion capture data (due to the uniqueness of the capture conditions), and results in synthesizing adaptative behaviors towards environment changes through the simulation of dynamic laws. Secondly, motion capture data allow the specification of the virtual character control by understanding the body characteristics and expressive components that are at stake during the motion. Moreover, the possible alteration of the original motion by the simulation process emphasizes the importance of a comprehensive and useable analysis of the original motion.

This paper presents the extraction of percussion grip parameters that can be used as gesture inputs for the dynamic control of virtual characters performing different technical percussion styles. It is organized as follows. Section 2 reviews previous work about physically-based character animation using motion capture data, style-based computer animation and the analysis of instrumental gestures. Section 3 introduces the overall context and the method used for extracting parameters that characterize percussion grips. The grip influence, in terms of end-point trajectories and angular constraints, is presented in section 4. Section 5 discusses the benefit of such parameters for our synthesis framework. Eventually, section 6 concludes with further perspectives.

2 Previous Work

Data-driven computer animation techniques have widely been studied in the past few years, taking advantage of the increasing availability of quantitative representations of the human motion thanks to the proliferation of acquisition systems.

Music-related contributions for virtual character animation often use a sound-driven approach, such as MIDI-driven [7] or performance-driven [15]. However, these finally prove a lack in taking into account the dynamic aspect of instrumental gestures.

The combination of a motion capture database and physically-based methods yields to frameworks that are well-suited for modeling human-like interactive bodies, since the capture of real movements and the simulation of dynamic laws intrinsically provide physically realistic and adaptative results. The use of physical models for dealing with motion capture data has been involved in motion editing, retargetting and generating transitions between motion clips ([11] [1]). Hybrid methods have also been explored, composing data-driven kinematic and dynamic controllers ([6] [14] [18]), based on the design of PD controllers ([8] [17]) which is also the approach of our work [3].

Regarding stylistic aspects of virtual character animation, previous works concentrate its efforts on extracting and generating expressive motion from motion capture data, and proposes methods based on signal processing ([4] [2]), statistical models ([16] [13]), or the optimization of physical models [9]. The robustness of these approaches however calls upon complex methods that move away, in a musical perspective, from a useable and comprehensible data analysis of the variability and the particularity of instrumental gestures. Such data analysis of instrumental gestures mostly rely on trajectory analysis [5] and classification/recognition methods of trajectory parameters ([10] [12]).

3 Method

The analysis of timpani performance is directed by the present methodology (Figure 1), taking advantage of a timpani performance database and yielding to the extraction of gesture parameters that can be used as gesture inputs for the control of physically-based humanoids.

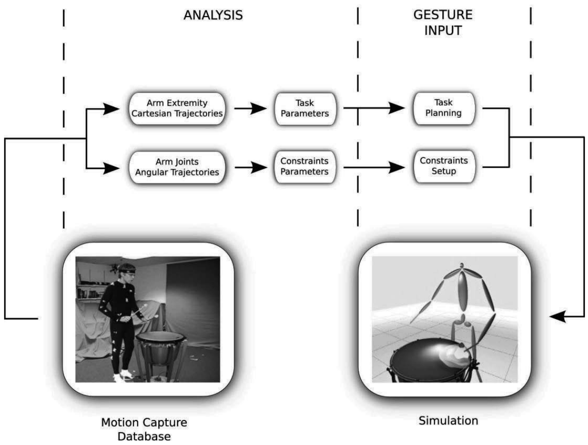


Figure 1: Methodology for the analysis of timpani performance.

A motion capture database was created, where three timpani players were asked to perform the capturing protocol designed for taking into account various timpani playing situations: drumstick grips, beat locations and musical varations. The motion was originally recorded using a Vicon 460 system with 6 cameras, sampled at 250 Hz. Markers have been placed both on the whole body of the performer (according to the Plug-in Gait model) and drumsticks. Each motion capture session results in the processing of the motion for a skeleton model (BVH format), including the position and orientation of the 19 joints composing the skeleton model.

The analysis of timpani instrumental gestures is achived by a two-level scheme considering task and constraint spaces. The task space can be defined as the 3D cartesian space containing end-arm and drumstick trajectories. The constraint space can be defined as the angular strategies that affect the physical components which condition the behavior of the performer, as angular trajectories and their derivatives can directly be correlated to joints' physical parameters such as stiffness or viscosity.

The parameterization of these spaces (task and constraint parameters) can be used as gesture inputs for the dynamic control of virtual characters [3] by task planning and constrain setup. In this paper, we focus especially on this parameterization step for studying the effect of drumstick grips on timpani instrumental gestures for the subjects S1 and S2.

4 Grip Effect: from the Tip of the Drumstick to Arm's Articulations

Variations in timpani playing range from drumstick grips, beat locations, to musical variations [3]. The present paper focuses specifically on the effect of french (F) and german (G) grips on timpani instrumental gestures, based on the analysis of beat attacks performed by two subjects S1 (French grip) and S2 (German grip). We focus on spatial and temporal differences that emerge from the analysis of the vertical component of the tip of the drumstick, as well as informations about arm articulations (shoulder, elbow, wrist) explaining these diffferences.

4.1 Axis Convention

Figure 2 defines the XYZ global coordinate system that has been used during motion capture sessions.

Every joint's position and orientation is later expressed towards this global frame. Special attention should be taken regarding the expression of joints' orientations. The geometric model issued from the motion capture software expresses orientations as XYZ euler sequences, the Z axis of an articulation being aligned by convention in the continuity of the next member. Especially, the rotations around X and Z axes will be respectively refered and denoted as flexion (rX) and twist (rZ) actions.

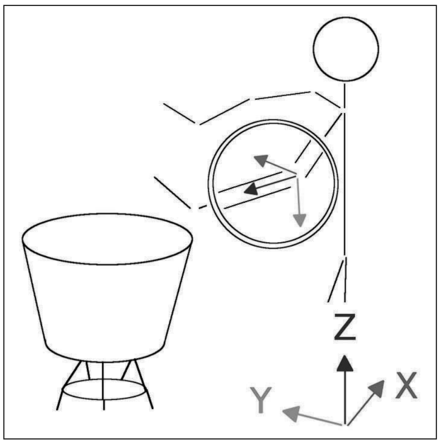


Figure 2: Experimental setup and axis convention.

4.2 Vertical Component of the Tip of the Drumstick

Quantitative features (Table 1) processed on the vertical component of the tip of the drumstick show that S1 (French grip) performs the same timpani gesture with much more amplitude. The mean of the height of the tip of the drumstick for the subject S1 is about twenty centimeters higher than for S2 (German grip), with a variance twice as high. This fact is strengthened by the vertical range of motion of the tip of the stick for S1 that is about twice as high than its counterpart for S2. The mean of the height of the tip of the drumstick shows moreover that the tip of the stick is in average closer to the timpani membrane for S2.

Table 1: Vertical component of the tip of the drumstick.

Subject (Grip)	S1 (F)	S2 (G)
Mean [mm]	1133	909
Variance [mm]	111.2	46.1
Range of Motion [mm]	434.2	178.6

Characteristic local extrema can also be observed during the preparatory gesture of beat attacks. Figure 3 presents an example of the vertical component of the preparatory gesture between two beat attacks, and the identification of three characteristic extrema denoted E1, E2 and E3. These extrema are temporally (temporal apparition in percentage of gesture's duration) and spatially characterized in Table 2.

Table 2: Local extrema of the vertical component of the tip of the drumstick: average temporal (in percentage of gesture's duration) and spatial characterization.

Subject (Grip)	S1 (F)	S2 (G)
E1 [% / mm]	21.2 / 1255.6	12.0 / 916.95
E2 [% / mm]	47.6 / 1059.9	68.3 / 881.58
E3 [% / mm]	68.4 / 1258.9	85.2 / 1022.4

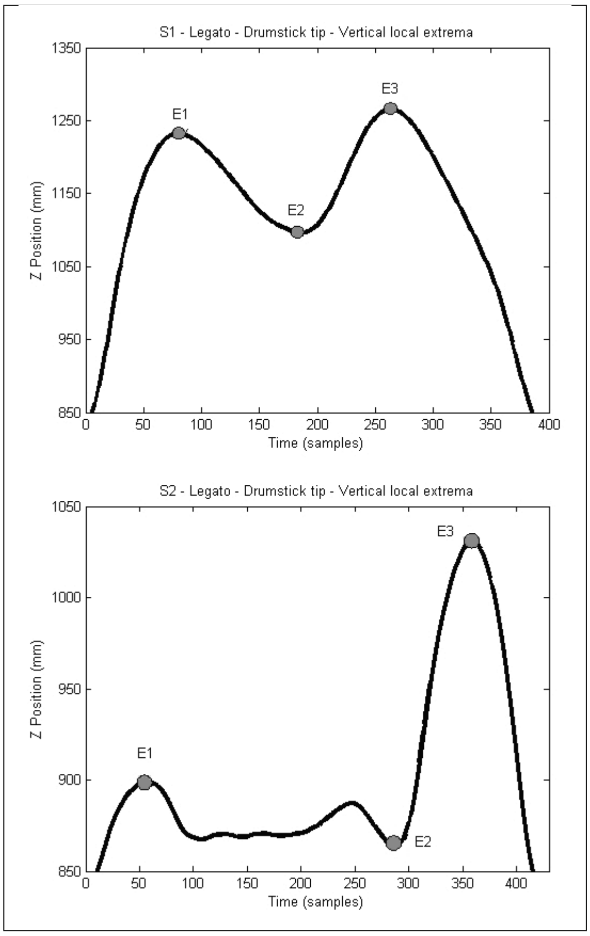


Figure 3: Local extrema of the vertical component of the tip of the drumstick for S1 (top) and S2 (bottom) between two beat attacks.

Vertical extrema E1, E2 and E3 (Table 2) are temporally equi-distributed for S1 showing a continuous preparatory gesture, whereas local extrema for S2 denote three discontinuous parts. E1 corresponds to the reaction of the previous rebound, between E1 and E2 the tip of the drumstick seems to seek a rest position (during more than the half of the whole movement duration) just above the timpani membrane, while E2 and E3 correspond to the amplitude that S2 gives for the next beat attack. Table 2 quantifies also the effect of the french and german grips on the vertical amplitude of the extrema.

French and german grips influence spatial and temporal characteristics of height of the tip of the drumsticks. The rest position between the extrema E1 and E2 for S2 could be explained by the proximity of the tip of the drumstick to the timpani membrane.

4.3 Shoulder Constraints

The analysis of the range of motion (Table 3) of shoulder's orientation and elbow's position shows that the upper-arm articulations of S2 are much more constrained than for S1. The range of motion of shoulder angles of S2 is so limited (between two and seven times lower than for S1) that the position of the elbow is almost constant. This could explain why the vertical amplitude of the tip of the drumstick for S2 is lower than for S1.

Table 3: Shoulder and elbow range of motion.

Subject (Grip)		S1 (F)	S2 (G)
Shoulder angle [°]	rX	27.00	8.42
	rY	21.81	2.97
	rZ	22.21	11.98
Elbow position [mm]	X	98.2	13.3
	Y	133.5	31.9
	Z	68.5	9.8

The stiffness of the upper-arm is influenced by the drumstick grip, and partly explains the difference in the amplitude of the drumstick. Elbow and wrist angle strategies also emphasize these differences.

4.4 Elbow and Wrist Constraints

Figure 4 shows the importance of the flexion of the elbow (S1) and the twist of the wrist (S2) angle trajectories. For both subjects, we can indeed correlate the temporal apparition of the extrema of these angles to the temporal apparition of the extrema quantified on the vertical component of the tip of the drumstick (Figure 3).

The relationship between elbow and wrist articulations for flexion and twist angle trajectories (Figure 4) shows morevover the trend of the wrist in amplifying different angles. For the subject S1, the flexion angle of the wrist is amplified towards the flexion angle of the elbow during the whole duration of the movement. Whereas for the subject S2, the twist angle appears more predominant as it is amplified towards the twist angle of the elbow during the whole duration of the movement.

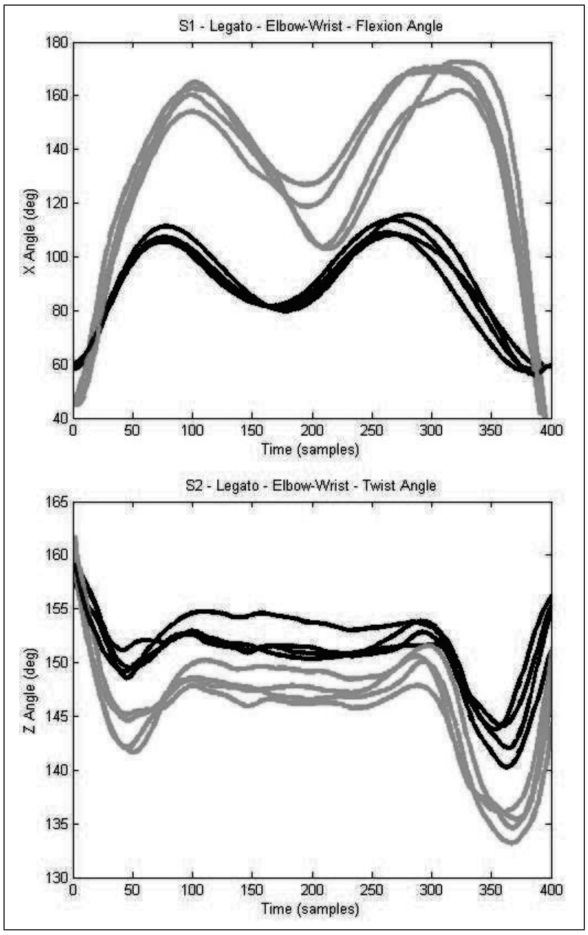


Figure 4: Intra-subject relationship between elbow (dark) and wrist (light) regarding flexion and twist angle trajectories for the subjects S1 (top) and S2 (bottom) between multiple beat attacks.

Due to this amplification relationship between elbow and wrist angles, we focus only on the elbow articulation for the flexion and twist angle trajectories for comparing S1 and S2 trajectories (Figure 5). For both flexion and twist angles, two different strategies can be put in evidence.

The flexion angle for S1 shows the same oscillatory motion as the drumstick, whereas S2 seems to lock the flexion angle to a low value for about the half of the movement, until the late preparation of the next beat attack.

As for the twist angle, two opposite strategies are used. On the one hand, for S2, the twist value at impact is very high, so that the interior of the forearm almost faces the timpani membrane. During the preparatory gesture, the twist angle is almost locked and the subject lately tends to gain momentum by decreasing this twist angle. On the other hand, S1 seems to less constrain the twist angle (attested by the variation during the preparatory gesture) and tends to gain momentum by increasing the twist angle between two beat attacks.

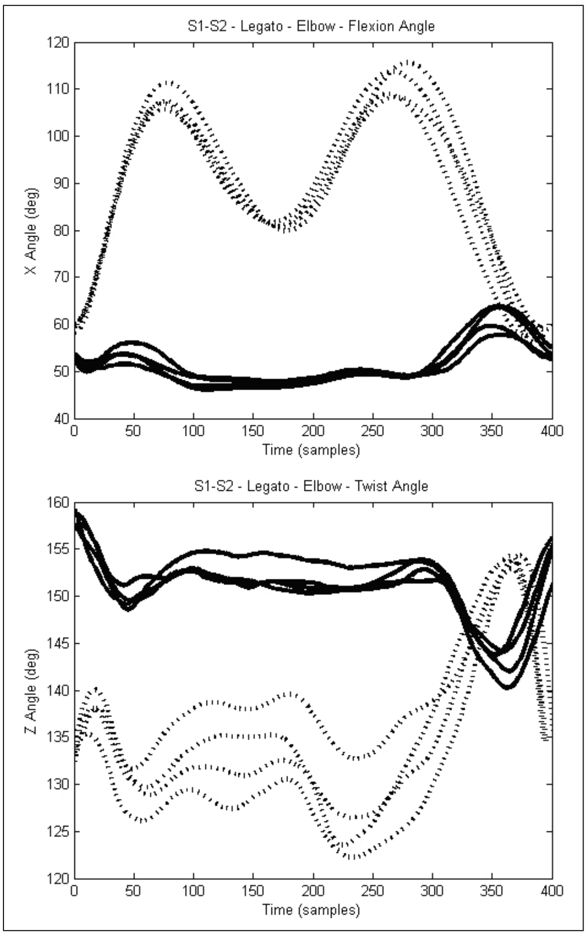


Figure 5: Inter-subject comparison of elbow flexion (top) and twist (bottom) angle trajectories for the subjects S1 (dash) and S2 (plain) between multiple beat attacks.

5 Benefit of Task and Constraint Parameterization for the Synthesis

Our system currently allows the simulation of virtual timpani instrumental situations, and the retrieval of kinematic and dynamic characteristics about timpani instrumental gestures (Figure 6). It takes directly angular trajectories as raw inputs from the motion capture database for the physical simulation of timpani instrumental gestures [3]. This input step thus does not discriminate between timpani playing styles that are inherent to performers.

The parameterization step described in section 4 is the first step towards such a discrimination and the modelling of timpani gestures, in terms of providing control parameters for upper-limb postures and the planning of gestures. The two subjects S1 and S2 under study have shown different strategies, both at the task and constraints levels, underlining parameters that can be handled by our system.

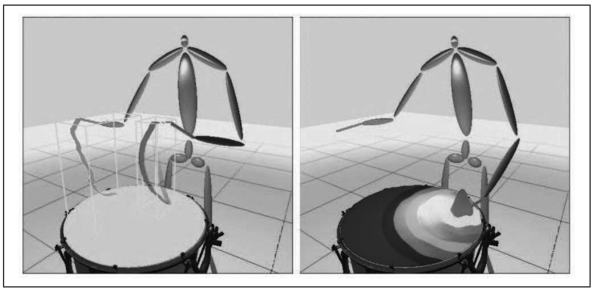


Figure 6: Retrieval of kinematic and dynamic motion cues thanks to the simulation framework of timpani instrumental gestures.

As for the task level, the local extrema that have been observed on the height of the tip of the drumstick are used as task control inputs to our system. Spatial and temporal characteristics yield the planning of timpani drumsticks motion during the simulation. As for the constraint level, the different strategies observed on the shoulder, elbow and wrist angles are used for the constraint setup of the degrees of freedom of virtual character joints. For S1, the oscillatory motion of the flexion angles for shoulder, elbow and wrist articulations is used for the simulation process. For S2, shoulder angles can be almost locked while the modelling and simulation of the elbow and wrist articulations involves twist angles.

6 Conclusion

We have presented in this paper the identification and the extraction of percussion technique parameters. The observation of differences on spatial and temporal characteristics of the height of the drumstick for two subjects with different grips has been explained by strategies involved in upper-arm stiffness, elbow and wrist angle trajectories. Such percussion style parameters can be used as inputs for our physically-based character animation framework for generating stylistic and expressive virtual timpani instrumental situations.

Future directions include the study of the effect of other timpani playing variations on the presented parameters, we aim namely at analysing the effect of beat locations and musical variations on timpani instrumental gestures.

7 Acknowledgments

The authors would like to thank Prof. Fabrice Marandola (McGill) and Prof. Sofia Dahl for helpful hints about percussion performance, as well as the timpani performers. This work is partially funded by the *Natural Sciences and Engineering Research Council of Canada* (Discovery and Special Research Opportunity grants), and the *Pôle de Compétitivité Bretagne Images & Réseaux*.

References

[1] Y. Abe, C. Liu, and Z. Popović. Momentum-based parameterization of dynamic character motion. In *Proc. of ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, 2004.

[2] K. Amaya, A. Bruderlin, and T. Calvert. Emotion from Motion. In *Proc. of the International Conference on Graphics Interface*, pages 222–229, 1996.

[3] A. Bouënard, S. Gibet, and M. M. Wanderley. Enhancing the Visualization of Percussion Gestures by Virtual Character Animation. In *Proc. of the International Conference on New Interfaces for Musical Expression*, pages 38–43, 2008.

[4] A. Bruderlin and L. Williams. Motion Signal Processing. In *Proc. of the Annual Conference on Computer Graphics and Interactive Techniques*, pages 97–104, 1995.

[5] S. Dahl. Playing the Accent: Comparing Striking Velocity and Timing in Ostinato Rhythm Performed by Four Drummers. *Acta Acoustica with Acoustica*, 90(4):762–776, 2004.

[6] P. Faloutsos, M. van de Panne, and D. Terzopoulos. Composable Controllers for Physics-Based Character Animation. In *Proc. of the Annual Conference on Computer Graphics and Interactive Techniques*, pages 251–260, 2001.

[7] R. Hanninen, L. Savioja, and T. Takala. Virtual concert performance - synthetic animated musicians playing in an acoustically simulated room. In *Proc. of the International Computer Music Conference*, pages 402–404, 1996.

[8] J. Hodgins, W. Wooten, D. Brogan, and J. O'Brien. Animating Human Athletics. In *Proc. of the Annual Conference on Computer Graphics and Interactive Techniques*, pages 71–78, 1995.

[9] C. Liu, A. Hertzmann, and Z. Popovic. Learning Physics-Based Motion Style with Nonlinear Inverse Optimization. *ACM Transactions on Graphics*, 24(3):1071–1081, 2005.

[10] C. Peiper, D. Warden, and G. Garnett. An Interface for Real-Time Classification of Articulations Produced by Violin Bowing. In *Proc. of the International Conference on New Interfaces for Musical Expression*, pages 192–197, 2003.

[11] Z. Popovic and A. Witkin. Physically Based Motion Transformation. In *Proc. of the Annual Conference on Computer Graphics and Interactive Techniques*, pages 11–20, 1999.

[12] N. Rasamimanana, E. Fléty, and F. Bevilacqua. Gesture Analysis of Violin Bow Strokes. *Gesture-Based Communication in Human-Computer Interaction*, LNAI 3881, Springer Verlag, pages 145-155, 2005.

[13] A. Shapiro, Y. Cao, and P. Faloutsos. Style Components. In *Proc. of the International Conference on Graphics Interface*, pages 33–39, 2006.

[14] A. Shapiro, F. Pighin, and P. Faloutsos. Hybrid Control For Interactive Character Animation. In *Proc. of the Pacific Conference on Computer Graphics and Applications*, pages 455–461, 2003.

[15] R. Taylor, D. Torres, and P. Boulanger. Using Music to Interact with a Virtual Character. In *Proc. of the International Conference on New Interfaces for Musical Expression*, pages 220–223, 2005.

[16] R. Urtasun, P. Glardon, R. Boulic, D. Thalmann, and P. Fua. Style-based Motion Synthesis. *Computer Graphics Forum*, 23(4):799–812, 2004.

[17] V. Zordan and J. Hodgins. Motion Capture-Driven Simulations that Hit and React. In *Proc. of ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 89–96, 2002.

[18] V. Zordan, A. Majkowska, B. Chiu, and M. Fast. Dynamic Response for Motion Capture Animation. *ACM Transactions on Graphics*, 24(3):697–701, 2005.

Human-Scale Bimanual Haptic Interface

Thomas Hulin Mikel Sagardia Jordi Artigas Simon Schaetzle Philipp Kremer Carsten Preusche
Institute of Robotics and Mechatronics
German Aerospace Center (DLR), Germany
E-mail: {firstname.lastname}@dlr.de

Abstract

This article presents a haptic system for bimanual haptic feedback that is composed of two light-weight robot arms. The system has a workspace and force capabilities similar to that of two human arms. Sophisticated control strategies are implemented to enable using this system as haptic interface. Depending on the requirements of the task one of three different handles can connect the human hand to the robot. Besides the human-machine-interface, the haptic rendering software is improved such that collisions of two dynamical interaction objects are computed in real-time. The system has been proven to be well suited as haptic interface for multimodal virtual assembly simulations.

1. Introduction

All systems designed to be operated by a human being have to provide an interface for their users, for example a keyboard and a monitor in the case of computers. This interface is called Human-Machine-Interface (HMI), sometimes also named as Human-System-Interface or Man-Machine-Interface. Systems built for applications like Virtual Reality (VR) simulations or teleoperation in which operators must be able to act intuitively require transparent HMIs, i.e. interfaces that display the virtual environment as realistically as possible.

The level of immersion in those environments depends strongly on the quality and quantity of feedback modalities (e.g. visual, acoustic, or haptic) that are provided by HMIs. For some applications like virtual assembly verification or remote maintenance, haptic feedback is crucial for task completion.

Applications like training of mechanics on virtual mock-ups, as required by the aeroplane industry, or assembly verifications for the automotive industry, call for workspaces and applicable forces roughly equal to those of human arms. Because of its specifications, the DLR Light-Weight Robot (LWR) [1] is well suited as kinesthetic-haptic device. One single LWR has a length of about one meter and is able to apply forces / torques of around $150\,N$ / $25\,Nm$ in any valid configuration.

Figure 1: Two-robot system for bimanual haptic feedback.

This paper describes a bimanual HMI, which is based upon two LWRs (Fig. 1). It is equipped with a head-mounted display for stereo-vision and stereo-acoustic feedback. The following section details the technical specifications of the haptic system and introduces three handles that connect the human hand to the LWRs. Section 3 discusses control issues for using the LWRs as haptic device. An evaluation of the system in two VR scenarios is described in section 4, while section 5 summarizes the main results and concludes with future work.

2. System Description

This article presents a haptic HMI for bimanual haptic feedback (shown in Fig. 1). The haptic system is

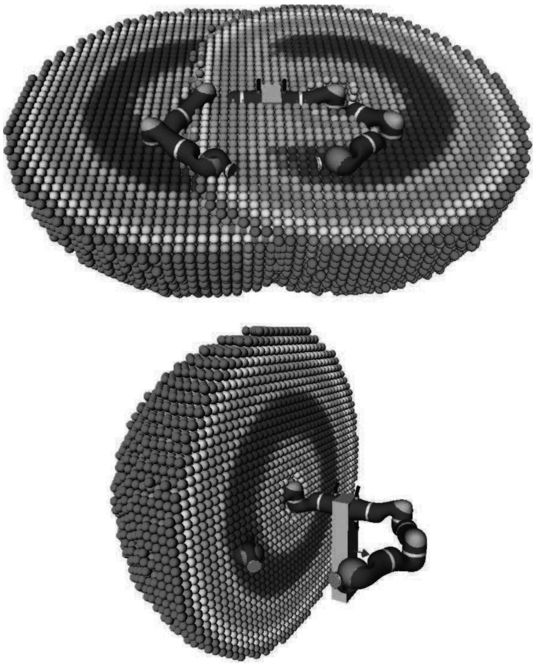


Figure 2: Sectional drawings of the workspace. Blue spheres mark points inside the workspace at which the robot can reach more than 75% of all possible three dimensional orientations, whereas red spheres mark points with less than 8% respectively.

Dynamic Mass	2 x $14\,kg$
Max. Payload	2 x $14\,kg$
Maximum Span	2 x $936\,mm$
Nr. of Joints	2 x 7
Sensors on each wrist	6-DoF Force-Torque Sensor
Sensors in each Joint	2 Position, 1 Torque Sensor
Sampling Rates	$40\,kHz$ current control $3\,kHz$ joint internal $1\,kHz$ cartesian
Motors	DLR-Robodrive
Gears	Harmonic Drive

Table 1: Specifications of the haptic system.

composed of two LWR arms [1] that are horizontally attached at a column.

Their workspace is similar to that of human arms. Two sectional drawings of the workspace are shown in Fig. 2. The spheres represent possible end-effector positions in the overlapping workspaces of the two LWRs. At each position, three dimensional orientations of the end-effector are checked for reachability. As result, sphere colors are indicating the reachability index at each point [2].

2.1. Light-Weight Robots (LWR)

The LWR is a light-weight, flexible, revolute joint robot, which by its overall sensory equipment is especially suited for working in the sensitive area of human interaction [1]. The robot's size, power and manipulation capabilities as well as its workspace are fairly similar to that of a human arm and they turn the LWR into a well suited HMI although it was not explicitly designed for this purpose. With its seven serially linked joints, the robot possesses a redundant kinematics that allows for null-space movement, which is valuable for avoiding collisions and optimizing the robot's configuration.

The LWR is equipped with very light gears, powerful motors and weight optimized brakes. As safety feature, these brakes need power supply to be released and they are activated as soon as the power is off. The electronics is integrated in each joint, including the power converters. The robot arms are able to handle loads up to $14\,kg$ in the whole workspace, while having a total weight of also $14\,kg$. Each of the LWRs joints has a motor position sensor and a sensor for joint position and joint torque. Thus, the robot can be operated for position, velocity and torque, being controlled at an update rate of $1\,kHz$, which allows for a highly dynamic behavior. An additional 6-Degree-of-Freedom (DoF) force-torque sensor is mounted on the wrist of each robot. This sensor can measure very precisely external forces, e.g. applied by a human operator.

For the use of the LWR as HMI, a main research topic focuses on safe human-robot interaction. Therefore, a biomechanical evaluation with crash tests has been carried out [3]. Furthermore, a thorough research on safety issues with respect to control strategies has been performed recently [4].

2.2. Handles for Haptic Interaction

A small manual flange allows changing fast the handle that connects the robot to the human hand. Three different handles are currently in use with the haptic system: a magnetic clutch, a grip-force interface and a joystick.

Magnetic clutch: the human hand is attached to a bracket in such a way that fingers are free to move (Fig. 3, left). Therefore, this interface can be used in combination with a tactile finger feedback device [5] or a finger-tracking device, e.g. the CyberGlove® [6], whose data can be used for visualizing a virtual hand in order to increase immersion, or even to control a multi-DoF device like the DLR Hand II [7]. This kind of hand attachment supersedes the visual tracking of the

Figure 3: Magnetic clutch (left), grip-force interface (middle) and joystick handle (right).

hand pose, because it can be calculated from the forward kinematics of the robot. The bracket itself is magnetically coupled to the robot flange. The geometry of the clutch, the arrangement of the magnets and their strength define the attaching forces and torques. If the applied forces or torques exceed this maximum force of the clutch, the user is detached from the robot and the integrated dead-man switch disconnects the power supply, which activates the brakes and stops immediately the robot.

Grip-force interface: a one-DoF grip-force feedback device with force feedback to the forefinger (Fig. 3, middle). This additional DoF can be used to grasp objects in a virtual reality simulation, or to explore their properties. Furthermore, this device can be used for telemanipulation tasks with force-feedback, e.g. closing a gripper with a certain force.

Joystick: a joystick handle equipped with a mini-joystick, a switch and several buttons, including a dead-man button (Fig. 3, right). This handle is especially suited for interactive tasks in virtual environments. Thus, the human user is able to change online control parameters of the robot, parameters of the virtual reality simulation, or adjust the visualization.

To increase immersion and to obtain a more intuitive impression of the virtual scenario, all the mentioned interfaces can be used in combination with a vibro-tactile feedback device [8] for haptic feedback to the forearm.

3. Control Issues

The control of the robot arms is challenging. On one side, due to its seven joints, the LWR has a redundant DoF that has to be controlled. On the other side, the robot's inertia must be scaled down in order to improve free-space movement behavior. And additionally, since the workspaces of the two robots overlap, collision detection and avoidance must be implemented.

This section assumes the robot being impedance controlled, i.e. torques are commanded to the robot and a backdrivable behavior is enabled in every joint.

3.1. Null-space Motion

As mentioned above, due to its seven joints, the LWR has a redundant DoF. This means that it can maintain a fixed pose with its end-effector, while moving freely its elbow. This redundant DoF can be used for two purposes: for optimizing the robot's configuration and as safety measure by featuring a compliant behavior.

The compliant behavior is inherent in the LWR, if it is impedance controlled, which makes possible pushing the elbow away.

The approach used for optimizing the configuration is described below. It allows for commanding a force at the elbow without disturbing the human operator. This desired force F_d can be set as parameter, for example such that it pushes the elbow of the right robot to the right, and vice versa for the left robot. Given a desired force, the corresponding desired torques T_d of the first four joints result as product of the transposed partial Jacobian matrix from the base (0) to the elbow (joint 4) with the six dimensional force vector F_d

$$T_d = {}^0J_4^T F_d. \tag{1}$$

The resulting force at the end-effector caused by T_d is determined by

$$F_E = ({}^0J_7^T)^+ \cdot \begin{bmatrix} T_d \\ --- \\ 0 \end{bmatrix}, \tag{2}$$

where $({}^0J_7^T)^+$ is the pseudoinverse of the transposed full Jacobian matrix. Given that the elbow motion should not have any effect on the end-effector, the force F_E must be compensated. The required joint torque T_{comp} to compensate the force F_E at the end-effector is

$$T_{comp} = -{}^0J_7^T \cdot F_E. \quad (3)$$

Therefore, for the commanded torques it yields

$$T = \begin{bmatrix} T_d \\ --- \\ 0 \end{bmatrix} + T_{comp}. \quad (4)$$

The torques T push the elbow in the direction of F_d, while not affecting the end-effector. Note that in singular configurations T vanishes.

Due to the fact of non-ideal actuators (friction, ripple, dynamics), the commanded torques do not correspond exactly to the performed torques, and the elbow motion is slightly perceivable.

3.2. Feedforward

As stated above, the mass of the each robot is of about $14\,kg$. Although the robots' gravitation is compensated, a human operator would soon get tired during a haptic simulation, due to the inertia. To avoid this, the perceived inertia must be reduced. This requires measuring the external forces at the end-effector.

As the robot's mass matrix is known, these forces can be determined from the torques of the seven joints. Yet, two problems arise with this approach. First, the calculated force at the end-effector is not accurate, because the model parameters of the robot are not perfect. And second, in singular configurations not all six values of the end-effector's force vector can be determined.

On account of this, a force-torque sensor is mounted on each robot's wrist. This sensor measures very accurately the applied forces, which can be used for scaling down the inertia of the robots.

For the LWR a feedforward compensation is applied, such as described in [9]. Different gains are used for translation and rotation, in order to obtain a stronger reduction of the inertia for rotations around the human's hand. With this feedforward compensation the perceived translational inertia is reduced to 33%, and the rotational inertia to 25% of their original values.

3.3. Collision detection

In contrast to many other haptic systems, which are mechanically designed such that collisions are impossible, the presented haptic system can collide. On the one hand, as described before, the workspaces of the two LWRs intersect. On the other hand, the robots can collide against the table on which they are mounted.

Although the robots are operated impedance controlled — they can be pushed away at each segment — a collision detection between the robotic arms is required in order to prevent damage of the robot arms.

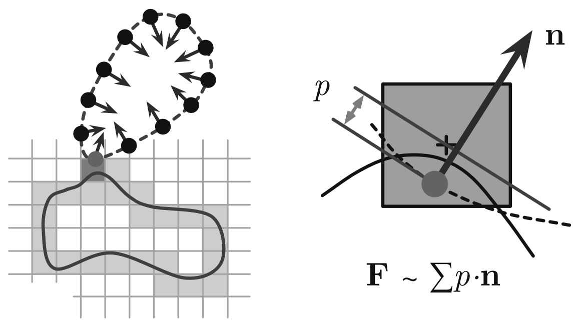


Figure 4: Schematic description of the Voxmap-PointShell® Algorithm. Left: voxmap and pointshell structures are colliding; right: single force computation in a colliding surface-voxel.

The implemented collision avoidance algorithm is based on the robot's partial forward kinematics, i.e. transformation matrices from the robot base to each joint, which specify the pose of each joint. As soon as the distance between a joint and the other robot or the table is below a certain radius, a spring-like force pushes against the imminent collision. In addition to that of the joints, collision avoidance for the three introduced handles is implemented in a similar way.

4. Virtual Reality (VR) Applications

The HMI is used with two different VR scenarios described in this section. For both, two virtual and moveable objects are coupled to the robots. By the use of a tracking system for the HMD, the operator is able to look around in the virtual environment. The virtual scene is rendered in the haptic and visual domain. Both are displayed to the human operator using the two robots of the HMI and the HMD respectively. Depending on the requirements of the task, one of the three introduced handles can be used for each robot.

This section is divided into two subsections: the first one explains the haptic rendering algorithm used to compute collisions and forces in the virtual world, whereas the second subsection describes the performed tests.

4.1. Voxmap-Pointshell® Algorithm

The haptic rendering algorithm used to compute collision forces and torques for the haptic feedback is an adapted version of the Voxmap-PointShell® (VPS) Algorithm [10]. The VPS algorithm enables haptic rendering of virtual objects moving in almost arbitrarily complex virtual environments at an update rate of $1\,kHz$.

Two data-structures are used to compute collision responses: *voxmaps* and *pointshells*, as shown in Fig. 4. Voxmaps are voxelized volume structures arranged for simulating the static properties of the objects. Pointshells, on the other side, represent dynamic or moving objects through clouds of points; each point is located on the surface of the polygonal model and possesses a normal vector pointing inwards the object.

Both data-structures are generated offline, and the algorithms to obtain them are explained in detail in [11]; a short description is given in the following lines, though.

The voxelizing algorithm navigates fast in the gridded bounding box of each triangle of the polygonal model and performs collision tests between candidate surface-voxels (Fig. 4 left, grey shadowed voxels) and triangles using the *Separating Axis Theorem* (SAT); if a voxel is colliding with a triangle, it is marked as a surface-voxel. The SAT simplifies the collision check problem to one dimension stating that two convex shapes do not intersect with each other if and only if no axis exists such that the projections of the shapes on it overlap. The axes to test are $\mathbf{a} \in \{\mathbf{e_i}, \mathbf{f_j}, \mathbf{e_i} \times \mathbf{f_j}, \forall i, j \in \{1,2,3\}\}$, being $\mathbf{e_i}$ the cartesian coordinate axes set in the center of the voxel and $\mathbf{f_j}$ the edges of the triangles. After obtaining all the surface-voxels, the voxmap is layered generating a distance-field inwards and outwards the object.

The pointshell generator uses the previously computed voxmap structure of the model and projects surface-voxel centers on the triangles. This is achieved minimizing the square distance function $Q = \|\mathbf{T} - c\|^2$, where $\mathbf{T}$ is a vectorial expression of the triangle and c the center of the surface-voxel to be projected. Once the points are obtained, the normals are computed analyzing the voxmap neighborhood of the point.

During the haptic simulation, collision detection and force computation are performed every $1\,ms$ in the original VPS algorithm [12], traversing all the pointshell-points that are in the scene. Every time a point is inside a surface-voxel, a collision is detected. The penetration is calculated measuring the distance from the point to the normal plane that goes through the center of the voxel. As shown in Fig. 4, during each haptic cycle, the penetration and the normal of colliding point k yield a single collision force $\mathbf{F_k} = p \cdot \mathbf{n}$, and all the collision forces summed together yield the total repulsion force. More detailed explanations concerning the collision detection and force computation are given in [12, 13].

4.2. Performed experiments

Regarding the system configuration of the robots, two dynamic or moving objects have been considered in the simulations, each one controlled by one of the robots. Every haptic cycle three pairs of models must be checked for collision: (i) the right object controlled by the right robot against the static scene, (ii) the left object controlled by the left robot against the scene, and (iii) the right object against the left object.

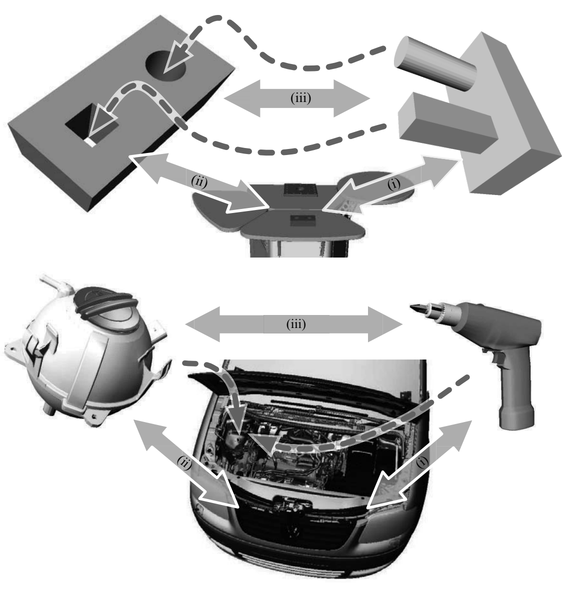


Figure 5: Simulations performed using VR. Top: *Peg-in-hole* test; bottom: assembly test. Dashed arrows represent movement paths, whereas solid arrows show the three pairs of objects to be checked for collision.

Two virtual reality simulations have been performed, both shown in Fig. 5. The first one consists of a *pins*-object coupled to the right robot and a *holes*-object moved by the left robot, constituting the classical *Peg-in-hole* benchmark set in a virtual static scenario containing a table.

The aim of this simulation is to show that the system stays stable during the simulation. In fact, although simple geometries are used, inserting a *pins*-object into a *holes*-object represents a challenging task in VR, given that typically it has been difficult to provide a stable and realistic haptic feedback coping with all the contacts that occur in the simulation. In addition, notice that in the particular moment where the *pins*-object is being inserted into the *holes*-object, a special situation comes up: the robots are coupled in such a way that their relative movement is only allowed in the direction established by the axis of the holes. This situation is handled successfully while maintaining stability through all the simulation.

The second simulation consists in assembling a coolant tank inside the engine hood of a VW Touran.

A remarkable feature of this simulation is the complexity of the objects in the virtual environment. The left robot is coupled with a virtual model of the coolant tank (25,263 triangles), while the other robot is coupled to a virtual electric drill (35,545 triangles). Both are within a scene occupied by the VW Touran model (3,364,266 triangles). Due to the large workspace of the robots, there is no need to scale the motion, i.e. moving a robot one meter will cause the corresponding virtual object to move one meter, too.

The goal of this scenario is to show that it is possible to check the suitability of virtual models in a very early design stage of product development, i.e. without building real mock-ups it can be verified whether the objects can be assembled and maintained. Therefore, possible designing errors can be easily detected and the development process of new cars can be sped up.

Moreover, since the system can be used intuitively, people that are not familiar with robots can also work with the system nearly without training. A practical example would be mechanics that can provide their knowledge and experience directly in the design process by checking assembly tasks on the virtual models.

5. Conclusions

This article introduced a human-scale system composed of two LWRs for bimanual haptic feedback. Control aspects like scaling down the robot's inertia, controlling the LWR's redundant degree of freedom for null-space motion, and avoiding collision were discussed.

With its workspace similar to that of both human arms, the high feedback forces and the control loop with an update rate of $1\,kHz$, this system suits very well as a haptic feedback interface, e.g. to perform virtual assembly tasks or to explore complex virtual scenarios without the need of scaling movements or forces. With the three described end-effectors the hardware setup can be adapted to the respective application.

A virtual assembly simulation of a VW Touran showed that even in complex scenarios the Voxmap-PointShell® Algorithm, which has been enhanced at the DLR [10, 11], is able to generate collision responses within $1\,ms$, and therefore it is suited as haptic rendering algorithm for the bimanual haptic system. Also classical benchmarks for haptic rendering — such as *Peg-in-hole*, in which a big number of contacts occur, without giving rise to high collision forces —, are stable with the system.

In future applications, this system may be used for telemanipulation, e.g. to control another two-robot system while feeding back haptic information.

Acknowledgment

The authors of this work would like to express their gratitude to the VRLab of the Volkswagen AG for supporting this research and for providing the models of the virtual car. This work has been also supported by the EU Government within the project "Multimodal Interfaces for Capturing and Transfer of Skill", project number FP6-IST-2005-035005-IP.

References

[1] G. Hirzinger, N. Sporer, A. Albu-Schäffer, M. Hähnle, R. Krenn, A. Pascucci, and M. Schedl. Dlr's torque-controlled light weight robot iii - are we reaching the technological limits now? In *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, volume 2, pages 1710–1716, Washington D.C., USA, 2002.

[2] F. Zacharias, Ch. Borst, and G. Hirzinger. Capturing robot workspace structure: representing robot capabilities. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 3229–3236, San Diego, California, Oct. 2007.

[3] S. Haddadin, A. Albu-Schäffer, and G. Hirzinger. Safe Physical Human-Robot Interaction: Measurements, Analysis & New Insights. In *International Symposium on Robotics Research (ISRR2007)*, pages 439–450, Hiroshima, Japan, Nov. 2007.

[4] S. Haddadin, A. Albu-Schäffer, A. De Luca, and G. Hirzinger. Collision Detection & Reaction: A Contribution to Safe Physical Human-Robot Interaction. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 3356–3363, Nice, France, Sep. 2008.

[5] T. Hulin, P. Kremer, R. Scheibe, S. Schätzle, and C. Preusche. Evaluating two novel tactile feedback devices. In *Enactive/07*, Grenoble, France, Nov. 2007.

[6] http://www.immersion.com, 2008.

[7] C. Borst, M. Fischer, S. Haidacher, H. Liu, and Hirzinger G. Dlr hand ii: Experiments and experiences with an anthropomorphic hand. In *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, Taipei, Taiwan, Sep. 2003.

[8] S. Schätzle, T. Hulin, C. Preusche, and G. Hirzinger. Evalution of vibro-tactile feedback to the human arm. In *EuroHaptics 2006*, pages 557–560, Paris, France, July 2006.

[9] J.J. Gil and E. Sánchez. Control algorithms for haptic interaction and modifying the dynamical behavior of the interface. In *2nd International Conference on Enactive Interfaces*, Genoa, Italy, Nov. 2005.

[10] M. Renz, C. Preusche, M. Pötke, H.-P. Kriegel, and G. Hirzinger. Stable Haptic Interaction with Virtual Environments using an Adapted Voxmap-Pointshell Algorithm. In *Proc. of Eurohaptics 2001*, Birmingham, UK, July 2001.

[11] M. Sagardia, T. Hulin, C. Preusche, and G. Hirzinger. Improvements of the voxmap-pointshell algorithm – fast generation of haptic data-structures. In *53rd Internationales Wissenschaftliches Kolloquium*, Ilmenau, Sep. 2008.

[12] W.A. McNeely, K.D. Puterbaugh, and J.J. Troy. Six Degree-of-Freedom Haptic Rendering Using Voxel Sampling. In *Proc. of SIGGRAPH*, pages 401–408, Los Angeles, California, Aug. 1999.

[13] W.A. McNeely, K.D. Puterbaugh, and J.J. Troy. Voxel-based 6-dof haptic rendering improvements. *Haptics-e*, 3(7), Jan. 2006.

Theory

Technical mediation of sensorimotor coupling : a minimalist approach

Charles Lenay,* Gunnar Declerck,* Olivier Gapenne,*
Barthelemy Maillet,* John Stewart,* & Indira Thouvenin**

(*)COSTECH, Université de Technologie de Compiègne, France
(**) HEUDIASYC, Université de Technologie de Compiègne, France

E-mail: charles.lenay@utc.fr

Abstract

The interface between an organism and its environment sets up the sensori-motor coupling between them; an interface can be termed "enactive" to the extent that it thereby participates in the active constitution of a perceived world and a point of view on that world. The scientific objective of studying the mediation of sensori-motor coupling by a technical device cannot be achieved by separating the components. It cannot be achieved by simply juxtaposing a psychophysiological study which separates the organism from its environment, a study of the physical properties of the real or simulated environment, and a study of the functional properties of the technical mediation. We propose here a method for directly studying the coupling as such. In order to illustrate a number of methodological principles, notably minimalism, we draw on several examples from the research programme developed at COSTECH (Compiègne University) over the last ten years.

1. Introduction

The general guiding principle for an enactive approach to perception is that the world of lived experience is constituted through the sensori-motor coupling between an organism and its environment. It follows that any modification of this coupling, in particular by employing technical interfaces, results in a transformation of the lived world. In order to understand how this coupling gives rise to a lived world, it would be vain to start by considering independently the psychology of an isolated subject on one hand and the physical environment on the other, and only subsequently studying their relations.

Indeed, a psychology which would attempt to study “natural” cognition separated from the external conditions of action and perception is condemned to have recourse to an internalist, representationalist view. In order to express the inappropriate character of a such a psychology, we may coin the pithy formula of Blaise Pascal (1623-1662) : « Descartes, useless and uncertain » [1]. For Pascal, the work of Descartes was « useless » because it did not take into account the practical aims of human action ; and « uncertain » because the foundational certainty of the *cogito* relies on an internal access and mastery of one’s own thought that does not really exist. In a similar vein, we may say that a representationalist psychology is uncertain (is it really possible to believe that the contents of our experience could be entirely produced by the internal states of the brain?); and above all that it is useless, because if it were really the case that cognition and perception consists solely of the activation and manipulation of internal representations, how could we understand that a new technical mediation can effect such a profound transformation of lived experience? At best, one would have to suppose that each new mediation leads to a reconstruction of the internal mental model of this mediation, resulting in a reconfiguration of the representations. But even this sort of clumsy “doubling up” would not enable us to *understand* how a technical medium could creatively augment our cognitive or perceptive activity.

By contrast, in an enactive approach, the dynamic coupling between an organism and its environment is primary. It is only subsequently that it is possible to distinguish the point of view of the subject on one hand, and the objects of their lived world on the other; and even then, viewpoint and objects remain fundamentally relational, and continue to mutually define each other

How then is it possible to conduct a scientific study which takes the coupling as its primary object, *before* separately defining the properties of the organism and its environment? The aim of this paper is to propose a research methodology which, we claim, is able to meet just this difficulty.

2. Methodological principles

2.1 The technical prism

The coupling between an organism and its environment is defined by two relations: « a », the repertoire of possible actions going from the organism towards the environment; and « s », the repertoire of accessible sensations going from the environment towards the organism (Fig.1). These repertoires depend on the coupling device in question: natural organs (biological motor and sensory systems), or artificial tools and interfaces that can be grasped by the organism.

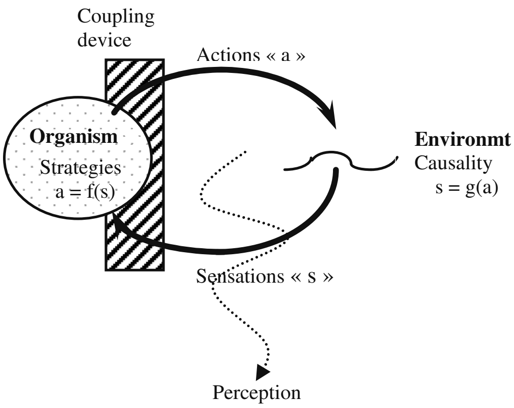


Figure 1. *Schematic diagram of sensorimotor coupling*

The technical mediation we propose to study here is both the object and at the same time the means of the study.

At the level of empirical analysis, the technical interface functions as a "prism" which decomposes the components of the coupling, and makes it possible to vary them experimentally. This is particularly true in the case of digital interfaces, which by their very nature allow a decomposition of the functions in discrete recombinable elements.

2.2 Minimalist principle

The first methodological principle we propose is that of *minimalism.* By reducing the sensory inputs *s* to a strict minimum, one can force an externalisation of the actions *a* which renders them readily observable. And by reducing also the actions *a* to a strict minimum, one renders feasible the ambition of having a *complete* observation of both *a* and *s* ; which in turn opens up the possibility of a complete modelling of both the environmental causality (a→s), and what we call the "strategy" (s→a) by which the organism generates a stream of actions as a function of the sensory feedback. Together, strategy and environmental causality generate a sensorimotor dynamics; the aim of the research is to identify the invariant sensori-motor laws which characterize this dynamics.

2.3 The principle of double description

An additional advantage of technical mediation is that it makes it possible to work with conscious adult subjects, and to study not only the dynamic coupling at a given moment but also the *genesis* of strategies which allow mastery of the dynamics. Thus, it is possible to ask the subjects to describe their own lived experience, and in particular the strategies they employ in order to identify the invariants of the sensori-motor dynamics in which they are engaged. The second principle of our approach is thus to systematically combine third-person empirical analysis with first-person phenomenological descriptions of the lived experience of the subjects. Here, it is possible to establish the correlations between the objective changes in the technical mediation, and the corresponding changes in the subjective constitution of a lived world of experience [2].

These two principles – minimalism and first-third person descriptions – will be briefly illustrated by work with a particular sort of interface: systems of "sensory substitution".

2.4 Perceptual supplementation

Sensory substitution systems consist of transforming information from one sensory modality (e.g. vision) into stimuli from another sensory modality (e.g. touch). Thus, in the classic TVSS ('Tactile Vision Substitution System', [3]), the image captured by a video camera is converted into a "tactile image" consisting of a 20 x 20 matrix of stimulators placed on the back, on the chest or more recently on the tongue.

The crucial lesson from this sort of system is that if the user remains passive, the stimulations are felt proximally at the surface of the skin, perception is poor and learning very slow and difficult. Everything changes, however, when the user himself is able to actively manipulate the camera. He begins by learning how the variations in his sensation are related to variations in his actions (a→s) : when he points the camera from left to right, the stimuli on his skin move from right to left; when he “zooms” forward, the points move apart; and so on. Under these active conditions, the user rapidly develops spectacular capacities to recognize shapes. He first learns how to point the camera at a target; he then discriminates lines and volumes; and finally recognizes familiar objects of increasing complexity, culminating in a capacity to recognize individual faces. Moreover, this capacity to recognize shapes is accompanied by an *exteriorisation* of percepts which become objects situated "out there" in a distal space. When this happens, the subject situates himself as a movable “viewpoint” with respect to the objects [4]. There are now many sensory substitution systems [5]. However, these systems are never an exact replacement of a missing modality; rather, they each give access to a *new* perceptual modality. We therefore prefer to speak of “perceptual supplementation” systems [6].

We consider that the extreme technical situation represented by these systems can serve as an exemplary paradigm. It is possible to have complete knowledge of the flow of sensory inputs and of the actions performed, throughout the period leading from the initial contact with the device up to the moment when objects are clearly perceived. It is only through this technical mediation that the user is able to constitute the new perceptual contents and perceptual space. The technical mediation “creates a possibility”, but it does not “determine” any particular content; the technology is enabling but not determining.

In general, the lived body of an organism defines the system of possible actions and sensations, and thereby defines the accessible perceptions. Prosthetic devices, like any tool that can be grasped, transform the possibilities of the lived body and open new fields of possible perceptions. Perceptual supplementation systems therefore allow a fundamental study of both perception in general, and of the role of technical mediations. On one hand, they provide an experimental access to important questions such as the consciousness of objective things in a space “out there”, because they make it possible to follow and to reproduce the genesis of this consciousness in the adult. On the other hand, by allowing systematic modifications in the capacities to act and to feel, they allow a precise analysis of the way in which technical mediations are constitutive of perceptual activity. Perceptual supplementation systems can play a paradigmatic role, because they allow for clear and complete control of the repertoires of sensations (the distributed sensory stimuli) and of actions (the possible movements of the receptor fields).

3. Perceptual space

The experiments performed by our group employ the ‘Tactos’ system which we have developped at the Technological University of Compiègne, in order to render images and graphics in a digital space accessible to blind subjects [7][8]. Tactos consists essentially of a set of tactile stimulators (a battery of two Braille cells with pins which can be activated electronically) which can be controlled by the movements of a set of receptor fields. An effector (stylus on a graphic tablette, computer mouse…) controls the movements of the receptor fields in the digital space (for example, a 4 x 4 matrix of 16 receptor fields in the plane of the computer screen). When a receptor field encounters at least one black pixel in the figure on the screen, this causes the all-or-nothing activation of the corresponding pin on the Braille cell (Fig. 2). This allows the exploration of a two-dimensional virtual tactile image.

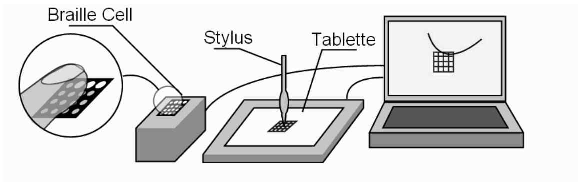


Figure 2. 2D Tactos system

Even in the case when the sensory input is reduced to a single tactile stimulator, corresponding to a single receptor field (which is indeed the strict minimum), a capacity to localize and to perceive shapes is conserved. The position of an immobile spatial singularity is constituted by the stable anticipation of the tactile stimulation depending on the movements of the stylus. It is by integrating the sequence of movements that the subject can succeed in constituting simultaneously the position of the shape and his own position with respect to the shape. The temporal succession of sensory stimulations is only perceived as a succession of contacts with an external form on the condition that the subject can *reversibly* come and go around the singularity that corresponds to the point of stimulation.

It is of course possible to reduce the movements of the receptor field even more by restricting them to a one-dimensional space. Here again, it is by a *reversible* exploration of this space that the subject will be able to localize the entities that he perceives as *distal stable objects,* whereas the sensory stimuli vary continuously as a function of his movements.

In any case, we can well appreciate here the essential role played by *action* in the progressive emergence of structured perceptions. What is perceived and recognized, is not so much the invariants of the sensations, but rather the invariants of the *sensory-motor cycles* that are inseparable from the activity of the subject [9][10][11][12]. If the form to be perceived is moving slowly, the subject will still be able to spatialize it by coming and going around it; he will then be able to perceive this movement by successfully anticipating the drift in his own rapid sweeping movements. However, if the form moves too fast, it will be lost and its spatialization becomes impossible.

4. Shape recognition

Even with only a single receptor field, blindfolded subjects are able to recognize shapes with the 2D Tactos system. A shape is not given to the sensory system all at once like a pattern applied to the skin. There is only one receptor field, and therefore only one sensation at any one point in time, so that the input signal has no intrinsic spatiality at all. If the subjects do nevertheless succeed in recognizing shapes, it can only be because of their active exploration, by integrating their movements and corresponding sensory feedbacks over time. Classically, the cognitive operations necessary for discriminating and categorizing shapes are considered as occurring mainly in the brain. Here however, thanks to the minimalist principle of reducing the sensory inputs to a single all-or-nothing signal, we have *forced a deployment of the perceptual activity in space and time*, which makes it easy to record this activity for subsequent analysis (see Figures 3a and 3b).

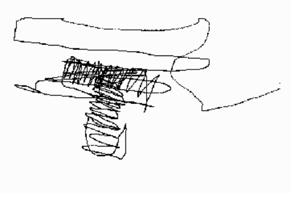


Figure 3a

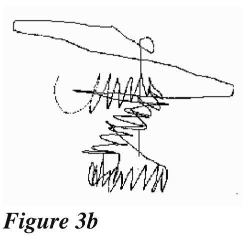


Figure 3b

These *perceptual trajectories* reveal several behavioural invariants. The subject starts by making wide exploratory movements; but as soon as a contact with the shape is obtained by crossing a line on the figure to be perceived, the subject rapidly converges on a strategy of *micro-sweeps,* oscillations of small magnitude around the source of stimulation. This is indeed an operation of localization. In these experiments, we observe that the imprecision of proprioceptive information [13] and/or memory limitations are such that if the subject moves too far away from the figure, he becomes lost and wanders around trying to recover the last point of contact. The subject does not seem to have access to the absolute position (in x-y co-ordinates) of the stylus on the tablette. He perceives only the direction and the amplitude of his movements, and even then with rather low precision. In these conditions, the immense advantage of the “micro-sweep” strategy is that it enables the subject to remain robustly in contact with the figure.

However, the micro-sweep gesture is not in itself sufficient to provide a perception of the figure. With unlimited memory and precise proprioception, we might envisage a strategy of “scanning” the figure, i.e. “mapping” the positions of all the points. The subject might then be able to calculate a “mental image” of the figure to be perceived. However, the incapacity of the subjects to identify the absolute positions in (x,y) renders this strategy ineffectual; and in fact it appears that the subjects do not proceed in this fashion.

The second characteristic of the perceptual trajectories therefore consists of a tangential displacement, following the local direction of the segment of the figure. Because of the limitations of the single receptor field, it is not possible to reliably follow the segment; inevitably, the subject ends up by drifting off the figure on one side or the other. If the subject had been attempting to follow the segment directly, the snag is that he then does not know on which *side* he has left the form, and consequently there is a high risk of becoming seriously lost. The strategy generally adopted, illustrated in Figures 3a and 3b, is thus a combination of the two elementary gestures: firstly, a rapid perpendicular oscillation which makes it possible to verify the position of the segment in this direction, and to “centre” the oscillations; secondly, a more gradual tangential movement which aims at following the whole length of the segment. This overall movement implements a second-order anticipation, based on a stability in the temporal frequency of the resultant stimulations.

This progressive exploration of the current segment is continually adjusted, taking into account possible surprises in the contour. Whenever contact with the figure is lost – for example, because of a sharp angle in the figure – the stylus backtracks, tries a new direction, and as soon as the new micro-sweep gives a regular stimulation again, continues on its way…. until the next detour. Each of these tangential directions is the confirmation – or the refutation – of an anticipation concerning the stimulations expected from the micro-sweep. We propose to consider that this stage corresponds to the recognition of “features” which together make up the perceived figure as a whole. Thus, the perceptual trajectories are both **recognition** and **enaction** of the figure. As in the phenomenological perspective, perception consists of succeeding (or failing!) to satisfy an active anticipation.

We could continue this analysis, going on to consider higher-order anticipations and enriched receptor fields. However, we will stop here in order to briefly mention another application of our minimalist methodology: the question of the constitution of a common world, and the recognition of other intentional subjects.

5. A common world and the recognition of Others

Here, the experimental device consists of connecting two Tactos systems to form a network. A single virtual digital space is shared via the network, so that the two blindfolded subjects can make tactile encounters with each other. For this, each subject must be provided with a "perceiving body" and an "image-body". Here again, the technical mediation plays the role of a prism, separating out components which are usually confounded. The "perceiving body" corresponds to the *Leib* of phenomenology, the body of lived first-person experience and the capacity to act and to feel. This is the body that is *constitutive* of experience and spatial objectivity. In this technical transposition, the « perceiving body » corresponds to

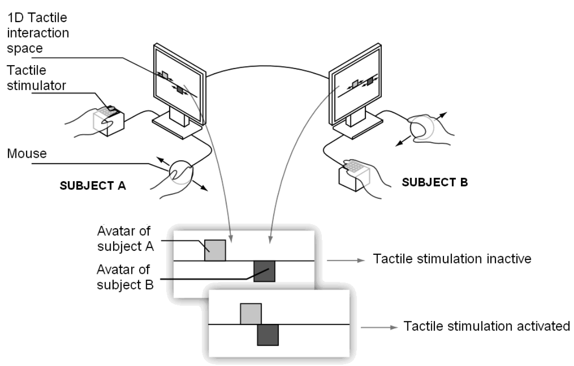


Figure 4. *Two Tactos systems in a network*

the receptor field that each subject can move in the virtual space. This "perceiving body" is to be distinguished from the "image-body", which is the body that the other subject can perceive. It corresponds to the *Körper* of phenomenology, the body that is *constituted* as a spatial object. In order for there to be a perceptual interaction, the perceiving-body of each subject must have attached to it an image-body that the other subject can perceive. The movements of the cursor of each subject thus controls both the movements of his perceiving-body, and those of his image-body that the other subject can perceive via his own receptor fields (Figure 4).

In order to simplify to the extreme the phenomenon of perceptual crossing, and in order to facilitate a precise analysis of the reciprocal dynamics of the interaction, the repertoire of sensations is reduced to a single all-or-nothing stimulation, as in the previous experiments; and the space of action is reduced to a one-dimensional world. Each subject explores this space of action, and encounters either passive fixed objects in this space, or the image-body of the other subject. It is remarkable that even in this (deliberately) impoverished situation, the subjects are able to distinguish their encounters with each other from encounters with passive objects including a moile lure [14]. The minimalism of this experimental setup makes it possible to make complete recordings of the perceptual trajectories of the two subjects; and the analysis of these trajectories allows us to formulate some hypotheses about the strategies employed by the subjects.

This analysis shows that if the subjects succeed in the task of recognizing each other, it is first and foremost because they succeed in situating themselves in front of each other. The discrimination of perceptual crossing, compared to other situations of stimulation, is based on an interdependence of the perceptual activities of the two subjects. The very content of their perceptual experience is formed by their essentially collective dynamics.

At the beginning of the experiment, during the period of familiarisation and learning, each subject is isolated in a one-dimensional perceptual space. The subject can perceive objects which produce a stimulation, but he does not perceive himself. His perceptual space is egocentric – but with a blind-spot concerning himself. Viewed from an outside third-person perspective, we can see that the perceptual spaces of the subjects are different. Framed from a first-person perspective, I do not perceive myself; and the image-body of the other appears to me as a simple object…… reciprocally, my partner does not perceive himself and *he* perceives *me* as a simple object. However, once our mutual recognition of the perceptual crossing has emerged – "catching each other's eye", as it would be in a visual context – the situation is altered. We now recognize each other as inhabiting a common space. The conditions for recognizing the presence of an Other are those for an object which is sensitive to my presence, a sensitivity linked to an activity oriented towards the perception of my presence, in other words a perceptual intentionality. The image-body of the Other therefore appears to me as being animated by an intentionality. The perceptual crossing reveals the living presence of an intentional activity.

From the point of view of the technical mediations, the interest of these experiments is to show that the recognition of the Other, with their intentional activity, does not necessarily result from an intermodal synthesis between signs of the presence of the Other which are already known, with novel phenomena. Here, the channel of interaction is perfectly controlled, and the subjects, who were naïve, had never used the system before. This mode of interaction is quite original, so it is impossible that the subjects were basing themselves on “innate” or previously acquired knowledge. It is therefore from the very inner workings of the novel perceptual modality employed by the interacting subjects, that this perception of the perceptual intentionality of the Other has been constituted.

6. Conclusion

In this paper, we have tried to sketch some of the analyses that are made possible by a minimalist approach. In each case, there are clearly many points which merit discussion, and many more experiments which could be done. But by trying to explain the interest of a few illustrative examples, we have hopefully shown that starting from the coupling itself, it is indeed quite feasible to understand the constitution of the corresponding lived experience.

When the laws of coupling make it possible to account for the constitution of space and objects, they correspond to the laws of physics. When the laws of coupling make it possible to account for the constitution of shapes and meaning, they correspond to the laws of psychology. And when the laws of coupling make it possible to account for the constitution of a common world and the institution of norms, they correspond to the laws of sociology.

References

[1] Pascal B. *Pensées.* Brunschvicg edition, Garnier, Paris, 1925.

[2] Husserl E. *Cartesian meditations.* Original 1931. Springer, 1977.

[3] Bach-y-Rita, P. ‘Sensory substitution in rehabilitation’, in L. Illis, M. Sedgwick & H. Granville (eds.) *Rehabilitation of the neurological patient*, pp. 361-83. Oxford: Blackwell Scientific Publications, 1982.

[4] Lenay, C., Canu, S. & Villon P. ‘Technology and perception : the contribution of sensory substitution systems’, in *Proceedings of the Second International Conference on Cognitive Technology,* pp. 44-53. Aizu, Japan, Los Alamitos , CA. (1997).

[5] Auvray, M. ‘Remplacer un sens par un autre : La suppléance perceptive’, in P. Fuchs, G. Moreau & J-.P. Papin (eds) *Le Traité de la réalité virtuelle III*, Vol.1, pp. 173-88. Paris : Les Presses de l’Ecole des mines. (2006).

[6] Lenay, C., Gapenne, O., Hanneton, S., Marque, C. & Genouel, C. ‘Sensory substitution : limits and perspectives’, in Y. Hatwell, A. Streri & E. Gentaz (eds) *Touching for knowing, cognitive psychology of haptic manual perception*. Amsterdam & Philadelphia: John Benjamins Publishing Company. (2003).

[7] Hanneton, S., Gapenne, O., Genouel, C., Lenay, C. & Marque, C. ‘Dynamics of shape recognition through a minimal visuo-tactile sensory substitution interface’, in *Third International Conference On Cognitive and Neural Systems*, pp. 26-29. Boston: Boston University Press. (1999).

[8] Sribunruangrit, N., Marque, C., Lenay, C., Gapenne, O. & Vanhoutte, C. ‘Speed-accuracy tradeoff during performance of a tracking task without visual feedback’, *IEEE transactions on neural systems and rehabilitation engineering* 12(1): 131-139. (2004).

[9] Piaget, J. *La naissance de l’intelligence chez l’enfant*. Neuchatel & Paris: Delachaux & Niestlé. (1936)

[10] Gibson, J.J. *The ecological approach to visual perception*. Hillsdale, NJ: Erlbaum (1986)

[11] Paillard, J. ‘Les déterminants moteurs de l'organisation de l'espace’, *Cahiers de psychologie* 14:261-316. (1971)

[12] Varela, F. *Principles of biological autonomy*. New York: Elsevier. (1979)

[13] Wann J.P. and Ibrahim S. ‘Does proprioception drift?’ *Experimental Brain Research* 91, 162-166, (1992).

[14] Auvray, M., Lenay, C. & Stewart, J. ‘Perceptual interactions in a minimalist virtual environment’, *New ideas in psychology* (in press) (2008).

Kinesthetic Thinking: Heightening Enactive Experience

Haakon Faste Elvira Todaro Massimo Bergamasco
PERCRO Perceptual Robotics Laboratory, Scuola Superiore Sant'Anna, Italy
E-mail: h.faste@sssup.it, e.todaro@sssup.it

Abstract

Humans are embodied beings—we think with our bodies, and our minds are embedded in the physicality of flesh. The challenge we face with enactive technology is that—as so eloquently examined in Thomas Nagel's classic essay "What is it Like to Be a Bat?"—consciousness makes mind-body problems fundamentally intractable [1]. No matter how intelligent an interface becomes, the user will always be of a different "mind." Research on enactive interfaces will nonetheless continue to push at this limit, requiring the design of new interaction modalities, such as kinesthetic thinking, that allow such synergies despite "divisions" of mind. The neurological and theoretical basis provided by this essay concludes that art, education, and design practice should play a central role in the implementation of such systems. Much further work in this area is needed.

1. Introduction

"Body and mind have evolved together throughout time and continue to act indissolubly, in time.
The ability to reason and solve problems is the result of a bodymind which is able to explore its own mental landscape and construct maps both within it and in the external world." —Andy Clark [2]

"The hand is action: it takes, creates, we could say that it thinks." —Henri Focillon [3]

Through the kinesthetic movement of virtual agents, enactive interfaces offer the possibility of extending embodiment from one mind to another. This presents a host of challenges to the designer of such systems, such as managing the experience of shared or co-embodied memories, and will require novel design strategies to be employed. We argue that comprehensive scientific advancements in this realm will require the development of the kind of intuitive and empathic thinking skills typically honed in the artistic context. It is our further contention that learning itself is inherently projected through the creative aspects of embodiment, which implies that emotional transfer through embodied experience could be greatly enhanced via open-ended and artistic enactive experiences. By offering a simulated experience of foreign bodies through cues which are far more than "simulations of skill," our aim would be to design enactive interfaces that enhance a user's sense of self.

1.1 Body Based Knowledge

We all have a body, through which we move and enter into contact with the world that surrounds us. Often, however, we lose "consciousness" of our body. This "way" of movement, our awareness of bodily manner and response, is largely regarded as non-integral to the emphasis of our culture. We move for many reasons, certainly not only to exercise, dance, or play sports.

Our bodies possess a secret language, the language of the body. Such "body language" is nonverbal and usually unconscious. It communicates through the use of postures, gestures, and facial expression. Moreover, it could be seen that our thoughts are the "gesture" of mind. Nowhere is this more evident than in the sharing of experience via enactive technology, a scenario that results from the complex structuring of space and time transferred from one body's "network" to another.

In *Cognition in the Wild*, anthropologist Edwin Hutchins performs a remarkable study of the spread of cognitive processes as they navigate across many brains, bodies, and machines, demonstrating human movement is a form of expression, interaction, and manipulation of material and social worlds [4]. The tactile and proprioceptive senses that guide our body's movement are a fundamental source of information concerning the work we perform, by which our bodies become culture.

In discussing the various stages of human development, the psychologist Jean Piaget is noted for his identification of three principal stages of growth: a body-based stage (in which children explore the world using predominantly kinesthetic senses), a visual stage, and the symbolic stage we associate with adult cognition [5]. Because each stage builds on the innate knowledge and wiring of the previous one, an increasing trend in philosophical and cognitive models of the human mind understands all linguistic and iconic knowledge in terms of "embodiment" [6, 7]. The study of the embodied mind is a rich unification of all academic fields, touching on computation, linguistics, developmental psychology, philosophy, religion, sport, craftsmanship, etc. [8, 9, 10]

1.2 Manual Skill

The internalized memory of specific task-based knowledge is what we refer to as manual skill. A particular skill is difficult to describe, as modeling a specific "act of doing" (albeit from a scientific perspective) is a continuing challenge for the fields of cognitive science, psychology, robotics, biomechanics and other behavior-related studies [11]. Research on the training of human skill represents a state-of-the-art issue in the field of laparoscopic and computer assisted surgery, for example, as well as in the maintenance activities in the industrial field where, at present, Augmented Reality technologies are used to assist the human operator in complex manipulative operations. A simplification of this process is attempted by capturing the forces involved by the performer of in such skills (as in the Figure, below). Current systems do not include capturing mechanisms that describe the special skills of highly qualified persons while executing their challenging manipulations (e.g. to describe what makes a good surgeon).

Because skills are learned through interaction with the world and with others, their definition is dependent on natural, social and cultural constraints. Skills range from functional movement, such as walking or running, to aesthetic and highly nuanced behaviors including sewing, welding, or ballet dance. In each of these cases, the memory and transmission of this information requires the demonstration of a teacher to a learner. Skill learning is thus an oral tradition, and has its roots in the history, needs, and "folklore" of a given region or industrial context. In terms of digital technologies, the acquired capability to handle specific skills with digital technologies can generate new ways of interaction with computers and communicate knowledge through it, making future software applications more accessible to non-specialist users and allowing the generation of digital archives of performed acts of doing which can then be replayed to a learning public.

New technologies enable new sources and methodologies for recording the present, and the conservation of the present is a fertile theme for research. The most recent generation of human-computer interface technology, including motion capture systems, brain computer interfaces, and haptic displays, return to humanity what previously had been taken away: the role of the body in manual work. Indeed, these technologies can conserve the memory of numerous occupations that, since the industrial revolution, have increasingly been on the verge of extinction; crafts such as woodcarving, shoemaking, and countless others. All of these are professions and techniques are still of great use, but are disappearing because they have no generational replacement. Looking back into history we find numerous examples of forgotten techniques that have quite current applications thanks to their futuristic aims (simply look at the machines of Leonardo, or techniques of the Renaissance such as perspective or printing). Society always seems to guide itself towards developing technologies rather than in other directions.

1.3 Embodied Emotion

There is more to embodied knowledge than muscle memory and proprioceptive learning. We are also able to think rational thoughts, retain memories, be creative, and empathize and express emotions. Today's understanding of human brain function shows it to be a complex mechanism for enabling these responses. At a high level its functions mirror their evolutionary development: in general, our instinctual behaviors are controlled by the brain stem, emotional behaviors by the limbic system, and rational behaviors by the cerebral cortex [12].

The instinctual brain, or brain stem, is concerned with self-preservation. It regulates automatic reactions of the body, such as blood circulation, muscle contraction, temperature regulation, sleeping, and breathing. Behaviorally it reacts to insure that the body is nurtured and nurtures others, mates, and defends itself ("fight"), escapes ("flight"), or submits to danger ("freeze"). The limbic system is the principal site of our emotions and memory storage, and operates as a "first alert" alarm system in times of stress and crisis. The rational brain is the site of our cognitive functions (thinking and consciousness), which are inhibited to lesser or greater degrees by these earlier systems during times of stress.

Body-based knowledge is thus less of an issue relating to the didactics of skill as it is a projection of thinking through movement. Our response to the stresses of modern life, including the use of "inadequate" enactive systems, demands that we modify how we interact with our bodies. For example, relaxation techniques allow for the cerebral cortex to function more optimally than it otherwise would. The Harvard Medical Mind Body Center suggests that individuals learn to relax by focusing on deep breathing, repetitive prayer, repetitive exercise, yoga, meditation, mindfulness, guided imagery, bodily awareness and progressive muscle relaxation [13]. Professionals working the fields of conflict management, education and psychology are increasingly aware that body-based knowledge is at the core of an individual's success. Paying attention to this knowledge is a central aspect to impacting positive social change.

2. Didactics, Creativity and Cognition in Enactive Interfaces

The instruction and learning of body-based knowledge is essential for the development of culture and the memory of skill. In particular, enactive interface technologies provide an exciting new platform by which body based information can be stored, shared, and experienced. Developing and preserving this knowledge, however, requires first-hand learning opportunities for students. Technology should support, not displace, non-technological efforts to educate bodies. Education has already made great strides in recognizing that each student's body has varying strengths and modalities of thought, including "multiple intelligences," one of which Gardner calls "bodily-kinesthetic thinking" [14]. University favoritism of research over creative fields such as body-based education and aesthetics is a disservice to society and a risk to survival [15, 16].

Increasing interest by culture in embodied learning is evident in the sheer number of online video courses on everything from dancing the tango to fixing a house. A growing desire for new teaching and learning mechanisms—especially for physical skills and activities—can also be observed in recent scientific literature. Basic human-computer interface tasks such as steering, aiming, or dragging all demonstrate the central importance of continuous body-based feedback and control. As the future of computing becomes increasingly multimodal, an emphasis on acquisition and transfer of physical activities will play an increasing role in the acquisition and transfer of skills in the virtual realm. Likewise, minds now unable to control bodies will become increasingly able. Just as desktop computers have enabled increased multi-tasking, virtual skills will be continuous, overlapping, and multifaceted. Given that today's estimates place the total number of computers worldwide at already over 1 billion computers, the future of learning will be increasingly embedded in digital systems. We should therefore comprehend our educational goals.

At a high level, a summary of the didactic process includes:

- A social obligation to: Transmit cultural knowledge (today understood not only as a pure and simple transmission, but to include appropriation, reworking, and production) in institutional and non-institutional forms; develop/educate the individual through learning experiences that become internalized and fuel further knowledge/abilities/expertise through the use of critical thought; develop specific sectors of knowledge/abilities/expertise in specific subjects when social institutions require it.
- Differing didactic processes depending on whether the individual has general, specific, or special objectives.
- Deliberate and projected acts of transmission/ communication, mediation and relationship.
- The three interrelated and coexisting periods of planning, action, and evaluation.

In this overview the importance of "didactic action" should be noted: it is conducted by the "teacher" on the basis of a plan; active participation of the subject is implied, and the real-time evaluation and adaptation of action is implicit in this context of interactivity; it takes place in an "atmosphere of learning" that connotes a relational type (climate, dialogue, narration, etc.) and a type of mediation (methodology, instrument, configuration, etc.) in which real-time evaluation and adaptation are integral. Just as there are strong parallels between computers and theatrical action [17], the corollary between embodied interaction and didactic action is an evident point of strength for enactive experiences.

3. Symbologies of Gesture

Discussing the role of gesture as an emotive quality for agent-based interaction, Laurel (1993) notes that "The representation of gestures by system-based agents can be constrained through the use of gestural qualities that derive from character traits and emotional states; for instance, the speed, force, and abruptness or percussiveness of gestures can be orchestrated to suggest such qualities as anger, assertiveness, gentleness, or lethargy." [17] Walter Benjamin once observed that the human body is small and fragile when confronted with "the mortal actions and explosions" it enables through our work. He defines work as a "place" where skill learning operates to simulate the interpretation of symbols communicated by bodily movements, allowing the understanding of culture through bodily action. Social history, in other words, reflects a tight relationship between the body and its work that allows it a series of significative "passages" [18].

The body, in fact, represents a regulating device at the center of knowledge, underlying the abilities of man. From a formal and structural viewpoint, the body defines a process of behavioral determination, centered (uniquely) on the specific body, even when it seems to be ignored and occluded, unthought of and hidden, by culture [19]. Bodily learning allows the body to remain itself—more similar, perhaps, to the body it learns from, but always a unique and ever modified being. Not by chance, the genealogical and hermeneutical exploration of a few segments regarding "knowings," the techniques, ideologies, practices and discourses relative to corporeality, provide interesting resolutions, precisely in the face of the history of work. Through the analysis of the treatment of bodies, it is possible to trace a more amplified and complex history, a truly definitive "social history of work," in which the body (and its emotional state) comes to assume a determinant role, particularly in the perceptive modalities and in the collective images that are stratified and articulated by it.

The interpretation of such symbols is carried out naturally within the culture to which work belongs; for other cultures, innovative technologies can give a real-time access to this kind of knowledge by means of enactive interfaces. Thus traditional forms of manual work are merged with technological research based on enactive concepts, addressing the increasing importance of performance and action in the process of contemporary craft. Furthermore, the emotive nature of gesture is a central component in establishing the "quality" of action. Enactive interfaces should thus seek to isolate interactions where emotion is articulated and recognized, rather than a consequence of expressing something through a transducer. We therefore use the definition of gesture articulated by Kurtenbach and Hulteen, that "A gesture is a motion of the body the contains information. Waving goodbye is a gesture." [20] Gestures can be classified according to their function as "semiotic", "ergotic" and "epistemic" [21], but we propose that the overarching gesture is the gesture of mind.

4. Towards Interface Environments for Heightened Embodiment

Enactive interface technologies offer a means to "extend" embodiment from one mind to another, through the kinesthetic movement of virtual agents. Through movement we pattern the memory of work, from "cellular memory" to the workings of "minds." Enactive interfaces can extend our environment to become interfaces of empathy. All enactive interfaces with this as a goal should therefore provide a targeted context for artistic learning. In the words of McLuhan (1968), "Anything that raises the environment to high intensity, whether it be a storm in nature or violent change resulting from new technology, turns the environment into an object of attention. When it becomes an object of attention, it assumes the character of an antienvironment or an art object." [22]

It is art that has lifted and continually rises to face the problem of relationships between emotion, perception, race, action, creativity, social instinct, cultural acquisition and so on; it is the nature of artistic form and its sensory aesthetics that efficiently synthesizes and gives solution to problems of social freedom. A cornerstone of artistic practice is embodied experience. Taken to their logical extreme, enactive systems will offer embodied experiences of fear, sadness and love.

The necessary cues endowed by an artist on his or her design—those that allow building such "bridges" into the user's mind—are what designers call "affordances" [23]. In enactive terms, they are embodied equivalents of the interactive "rollover" or "tooltip" metaphor: real-time context sensitive experience triggered only when the user's actions reflect a situation of need. Moreover, *the context of learning itself is inherently projected through the nature of artistic embodiment*. In traditional fields such as dance, music, or painting, the quality attributed to "artistic" activity (i.e., the "art of sailing"), and compositions involving technology and embodied experience, we put faith in our artists to endow meaning to others. By offering a simulated experience of foreign bodies through cues which are far more than the "simulations of skill," enactive interfaces may be designed to enhance a body's "sense of self."

The positive effects of multimodal feedback on human-computer interaction for embodied learning have been demonstrated by much recent scientific research. Such feedback spans the senses and includes audio, visual, and haptic/kinematic interaction [24, 25, 26]. Nevertheless, technologies of embodiment transfer will be hindered in the future by traditional screen-based mouse/pointer paradigms, mostly due to the low cost and pervasive use of such technologies in society today. To have a serious impact on the future of human-computer interaction, enactive systems must develop strategies to overcome these barriers [27]. The challenge here is that which has been faced by art throughout history: the desire of scientists to quantify what are essentially subjective measures.

Many of the physical actions and applications being analyzed by researchers in the fields of robotics and human cognition are relevant as new modalities for artistic embodiment. Intuitive new gestural interfaces, for example, will draw metaphors from these activities and integrate them in novel ways [28]. For this reason, the manner in which a user approaches objects and interacts with them through reaching, grasping, pushing, pulling, etc. should be analyzed by an intelligent system *with emotional empathy*. We must not only recognize the intent of a gesture, but also create predictive systems: *the anticipation of desire*. This level of empathy presents complicated issues for intelligence research, particularly in the realm of contextually embodied information. Like existing user interfaces, they will be heavily context dependent and involve subtle nuances depending on the task. Much like the embodied nature of non-digital skills, however, capturing and remembering these new behaviors as they are learned will prove to be a prudent and beneficial activity in the long term.

It is therefore important to pursue new research into the recognition and development of novel paradigms for embodied forms of kinesthetic thinking, ones which heighten the "relaxation of experience." A primary first step in this direction is to prototype and test intelligent emotion-aware enactive systems capable of analyzing continuous data in real time and to incorporate subjective feedback emphasizing bodily knowledge and proprioceptive awareness. Perhaps the process of emotional learning can be best addressed by exploring beyond the limitations of discrete applications and domains and focusing instead on the gesture *potential* of natural movement and the expressions capable through uninhibited gesture, particularly in intelligent and/or networked multimodal environments.

Finally, regardless of the technology used, a purely technology driven approach often fails to address fundamental human needs in interactions with digital information systems. In recent years, a methodology of human centered design has become central to state-of-the-art in human-computer interaction. This approach, applicable the observation of traditional skill learning environments as well as novel ones, has become popular in the commercial sector for its efficacy at revealing latent human need. Designers use subjective insight into human behavior to develop principles and strategies for inspired design. The power of this approach is its applicability and relevance across industries, disciplines, and technological platforms [29]. The desired result requires striking a balance between conventional elements and innovative interactions that enable an elegant usability experience. By insuring that the design is driven by an understanding of underlying human need, a design will be implemented that not only satisfies usability requirements but which is also both useful and delightful.

In the future, improvements to the kind of reactive virtual environment described in this paper will be developed and tested in the relevant fields of scientific research. Given the predictions of experts today, the efficiency of gesture learning on a variety of different users and situations is likely to have widespread implications for the future of work. For interactive learning environments in particular, performed data could be mined by an intelligent system to understand which aspects of movement are similar and how students' gestures evolve as they learn. Opportunities incorporating real time machine learning and interactive human/system interface didactics will be fascinating to apply in a networked environment where a variety of users collaborate to teach the system—and each other—the fundamentals of newly invented and/or improvised collaborative learning body.

References

[1] Thomas Nagel, *What is it Like to Be a Bat?*, in The Philosophical Review LXXXIII, 4, October 1974, pp. 435-50.

[2] Andy Clark, *Being There: putting brain, body, and world together again*, MIT Press, Cambridge MA, 1997, p.180

[3] Henri Focillon, *The Life of Forms in Art*, Zone Books, 1992

[4] Edward Hutchins, *Cognition in the Wild*, MIT Press, 1996

[5] B. Inhelder and Jean Piaget, *The Growth of Logical Thinking from Childhood to Adolescence*, Basic Books, New York, 1958

[6] Francesco Valera, Evan Thompson and Eleanor Rosch, *The Embodied Mind: Cognitive Science and Human Experience*, MIT Press, 1991.

[7] George Lakoff and Mark Johnson, *Philosophy in the Flesh: The embodied mind and its challenge to western thought*, Basic Books, New York, 1999.

[8] Andy Clark and David Chalmers, 'The Extended Mind' *Analysis 58* (1998), 7-19;

[9] John Haugeland, *Having Thought*, Harvard University Press, 1998

[10] Mark Rowlands, *The Body in Mind: understanding cognitive processes*, Cambridge University Press, 1999

[11] See, the European integrative research project "Skills," which deals with the acquisition, interpretation, storing and transfer of human skill by means of multimodal interfaces, robotics and virtual environment technologies. Information available at http://www.skills-ip.eu/index.php?en/46/overview, viewed May 19, 2008.

[12] Paul D. MacLean, *The Triune Brain in Evolution: Role in Paleocerebral Functions*, Plenum Press, New York, 1990

[13] Herbert Benson, *The Relaxation Response*, Harper Paperbacks, expanded updated edition, 2002

[14] Howard Gardner, *Frames of Mind: The Theory of Multiple Intelligences*, 1983

[15] Rolf A. Faste, "Ambidextrous Thinking," *Innovations in Mechanical Engineering Curricula for the 1990's*, American Society of Mechanical Engineers, 1994

[16] Rolf A. Faste, "The Role of Aesthetics in Engineering," Proc. *Japanese Society of Mechanical Engineering*, 1995

[17] Brenda Laurel, *Computers as Theater*, Addison-Wesley Professional, 1993.

[18] Walter Benjamin, *The Arcades Project*, (Belknap Press Edition 2002)

[19] F. Cambi and S. Ulivieri, *Storia dell'Infanzia nell'Italia Liberale (Childhood History in Liberal Italy)*, our translation, 1988.

[20] Gordon Kurtenbach and Eric A. Hulteen, *Gestures in Human-Computer Interaction*, The Art of Human-Computer Interface Design, Addison Wesley, 1990.

[21] Claude Cadoz, *Le geste, canal de communication home/machine*, Technique et science informatiques, 13, 1, pp. 31-61, 1994.

[22] Marshall McLuhan, *Through the Vanishing Point*, 1968

[23] Donald Norman, *The Design of Everyday Things*, Currency Books, 1990

[24] A. Kohlrausch and S. van der Par, "Audio-visual interaction in the context of multimedia applications," in *Communication Acoustics* (J. Blauert, ed.), pp. 109–138, Berlin, Germany: Springer Verlag, 2005.

[25] Carlo Alberto Avizzano and Massimo Bergamasco, *Haptic Interfaces: a New Interaction Paradigm*, Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems, 1999

[26] Jeff Lieberman and Cynthia Breazeal, *Development of a Wearable Vibrotactile Feedback Suit for Accelerated Human Motor Learning*, IEEE International Conference on Robotics and Automation, Roma, Italy, 2007.

[27] Elisabetta Sani, Emanuele Ruffaldi, Haakon Faste, Massimo Bergamasco, *Business Models for Enactive Interfaces*, Proc. Enactive/07, 2007

[28] Elvira Todaro, Haakon Faste, Otniel Portillo-Rodriguez, and Massimo Bergamasco, "The Hand Dance: A Didactic Performance Platform," in proc. *ICHIM International Cultural Heritage Informatics Meeting*, J. Trant and D. Bearman (eds). Toronto: Archives & Museum Informatics. 2007.

[29] Brenda Laurel (ed.), *Design Research: Methods and Perspectives*, MIT Press, 2003

Embodied, Disembodied and Re-embodied Cognition: the Potentials and Obstacles for Elegant Human–Computer Interfaces

Rachel Zahn
Université Paris 1 Panthéon Sorbonne, France
CREA, Ecole Polytechnique, France
E-mail: rachel.zahn@gmail.com

Abstract

It is clear that the ingenuity of brain interface technology will create significant changes in human lifestyles of the 21st century. Applications such as haptics, virtual reality, mind-reading computers, and perceptive pixel screens are pure genius. One can easily foresee the impact these developments might have on future generations.

Concurrently, a close intellectual neighbour, Enactive Cognitive Science, has been studying the dynamics of change in living systems. The enactive view of consciousness was introduced by the late Francisco Varela at CREA/LENA laboratories (Paris), one author of The Embodied Mind: Cognitive Science and Human Experience [1].

This paper will briefly explain Varela's neuro-phenomenological model for laboratory experiments, which could easily find a place within brain interface technology. We will also propose an unusual adaptation of this model: by working with specialists with exceptional psychophysical expertise in high proprioceptive performance, engineers can come closer to the truly elegant human-computer interfaces they seek.

1. Introduction

A successful technological future will require collaboration between researchers who enthusiastically explore disciplines as widely separated as the philosophy of mind and high performance training. Extradisciplinary input can often shed unusual and sometimes uncanny light upon mysteries that the constraints of one's own discipline cannot articulate. The inherent skills of these researchers must include a certain tolerance for *otherness* and a willingness to construct a new, interdisciplinary vocabulary capable of expressing what enactive cognitive science calls *neurophenomenological* evidence. A truly coherent analysis of enactive systems must cast its net of observation over multiple logical levels, widening its scope until a suitable embodied vocabulary can be constructed.

Francisco Varela insisted that laboratory science must develop methods for experimenting with and validating *awareness* of our *lived experience*. He modelled embodied cognition under laboratory conditions which required input from "first-, second- and third-person" sources. In this vein, he introduced a phenomenological vocabulary in order to better express the subtle relationships between the first person *subject* (e.g. the lived experience of a physically disabled person or a high performance artist) and the researcher (a laboratory engineer or technician with her third person scientific constraints). He even went so far as to introduce the "second person" (e.g. a trainer, coach, psychophysical specialist) into this interface as a significant catalyst in the experimental process [2].

This second person is the primary focus of this paper. Ideally, he or she is a *psychophysical specialist* in physical rehabilitation and/or highly skilled in proprioceptive performance. This specialist would study the embodied dynamics of the specific interface area being researched, and draw attention to critical moments in the psychophysical process of which the scientist has no working knowledge.

Such an intermediary will introduce a broader vision of the experiment and maintain focus on improving the internal psycho-sensory coordination of the subject, thereby improving the quality of performance achieved in the experiment. The expertise brought by the second person provides an unusual opportunity for the researcher to document the precise improvement of the subject's embodied *use* in relation to their manipulation of or by the technology.

2. Embodied cognition

"The emergence of embodied dynamicism in the 1990's coincided with a revival of scientific and philosophical interest in consciousness, together with a willingness to address the explanatory gap between scientific accounts of cognitive processes and human subjectivity and experience." [3].

In particular, the enactive approach of Varela, Thompson, and Rosch [1] introduced a new definition of cognition as "…the exercise of skilful know-how in situated and embodied action". This quality of being *situated* is an essential ingredient: the subject and scientist are seen as existing within the "conditions of possibility" [8] of living organisms, who are in the process of using an *as if* proposition of third person objectivity to create falsifiable laboratory data.

"The blind spot in the cognition sciences of the 20th century is that we do not have a method of properly accessing experience…

To deprive our scientific examination of this phenomenal realm amounts to…amputating human life of its most intimate domains, or else denying science explanatory access to it." [3].

Figure 1. Francisco Varela (1946-2001)

Their embodied enactive model has affected many related disciplines. For example, Thelen and Smith [4] applied dynamic systems theory to the enactive development of proprioceptive intelligence in children. It is very clear that the vestibular system is deeply connected to brain development before and after birth [5]. Today in biology and neuroscience, we find a recent and very strong interest in what is now called "the lost sixth sense" [6,7], the collaborative triad of vestibular, visual, and proprioceptive systems. Simply put, it means knowing where you are kinesthetically in relation to your most constant environmental influence: gravity. Researchers look for intra-species clues in evolution, in embryology, and throughout early childhood development. In this light, the quest of an *enactive* oriented scientist is to discover the *autopoietic* (self-organising) [8] logic that brought humans to upright bipedal posture. No one needs to teach a child to crawl, scoot, sit, and find its way to bipedal existence. This is no small task, as anyone who has ever been physically disabled knows. Interestingly, the most sophisticated psychophysical specialists show patients how to *re-access* rather than re-learn this autopoietic know-how during rehabilitation. These methods are non-coincidentally the same methods used in high performance training. The goal of both is *psychophysical congruence*, especially when adding a new skill such as the pole vault or playing a musical instrument.

Studying proprioceptive awareness in infants reveals a process so profoundly intelligent that when we have seen it, there can be no doubt that cognition is embodied from the very beginning. In fact, a great deal of this research on the "lost sixth sense" is driven by the practical problem of proprioceptive and vestibular disorientation in micro-gravity, which can leave astronauts as helpless as infants. Equally applicable to NASA research are cases like "Ian Waterman who, at the age of nineteen, lost the senses of touch and movement and position sense, relearned, over 2 years, to move by thought and by visual supervision." [9]. The sense of *agency* which a child develops by learning to lift and orient its head turns out to be a key element in both evolution and the ability to develop more refined skills. The loss of this *know-how* is a loss of self. By ignoring the significance of this basic sense of self, we humans have managed to *lose* touch with an essential intelligence. We have replaced it with a pretense: the disembodied intellectual agent. Enactive cognitive science attempts to correct this mistake. How this disembodiment comes about and how it may be resolved is also the focus of psychophysical specialists.

3. Disembodied cognition

Enactive *disembodiment* of mind is often overlooked by philosophers and even neuroscientists. It is a very difficult subject to approach because it challenges the *beholder* in the same way that the child in the story "The emperor has no clothes" strips cherished *as-if* beliefs from the powerful.

As the reader may remember, there is a traditional discussion in the philosophy of science which is called the *mind–body problem.* It is usually attributed to the writings of René Descartes in the seventeenth century, though strains of this problem can be traced back to the early Greek philosophers. In 1623, Descartes dared to put all that he thought to be real in doubt and even went so far as to question his own existence. Unfortunately, his conclusions about separate but interacting mind and body and his quest for *certainty* in a yet undiscovered quantum world have proved to be in error. The side effects of these errors are not easily undone. They manifest as mind-body dissociation, and in the worst case as *dissociative disorder*, a serious psychological pathology.

How could this come about? It turns out that the initial signs of dissociation can occur during a child's early developmental cycle or first school years. According to Daniel Stern, children develop their core sense of "self agency, self coherence, and self affectivity" between two and nine months [10]. In the 1930's, Myrtle McGraw observed that self-consciousness and self-confidence develop along with the proprioceptive mastery of stable bipedal coordination. [11]. Lise Elliot describes several developmental benchmarks: 2 years for an awareness of "I", "Me" and "Mine"; 3 years old for the beginnings of cognitive differentiation between *reality* and *appearance*; 3-5 years for a "theory of mind" capable of distinguishing memory, dreams, desires, beliefs, and imagination; and 6 years for universal brain maturation manifesting as skilful drawing, language, memory, attention, control, and self-awareness [5]. During these tender cognitive ages, children are spontaneous learners with predominantly neuroplastic brains. And it is precisely then that the greatest potential for dissociation occurs. At six years old if not earlier, children are expected to spend many hours a day in school, many days a year.

For the most part, they are taught to transfer their allegiance from first person lived experience to third person validation of their worth. In many cases, this induction into Cartesian duality happens at far too young an age. It creates a profound disruption in the newly formed *self* that arose from psychophysical know-how about horizontal and vertical negotiation of gravity. It is this dissociation from first person know-how that makes it particularly difficult for adults to re-learn walking or even to begin to move an arm after a stroke.

The lingering treatment of the body as a third person object, a non-reasoning, irrelevant but necessary condition of existence is the result of mind-body dualism embedded in our culture, education and beliefs about mind. It lingers like a resistant artifact even as new research completely invalidates its premise. For example, the highly publicised results of *mirror neurons* [12] and *brain neuroplasticity* [13] have severely upset previous beliefs about *closed systems* on the one hand and the *unchangeability of the mature brain* on the other. Not only has quantum mechanics shown the observer to be an essential *participant* in any system, we now think that mirror neurons are constantly influencing and being influenced at an interactive level of neurological simulation. Therefore, every psychophysical attitude and incident of communication is immediately mirrored and/or simulated by the receiving brain—regardless of whether a laboratory technician thinks she is in an objective (disembodied) state of observation. There is no doubt that third person scientific constraints are useful. It is the side effects that concern us deeply.

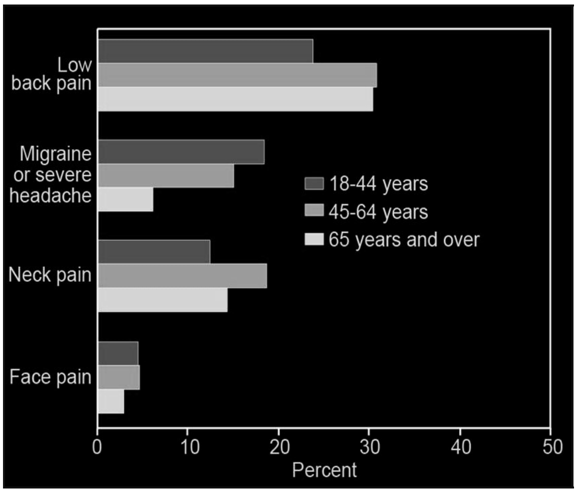


Figure 2. Lower back pain, neck, migraines and face pain, in the past 3 months (2004)

In Figure 2 we can see examples of these side effects, which we consider in most cases to directly result from the patient's lost sense of agency. They only notice the trouble when it has reached chronic proportions. Psychophysical experts have watched the proprioceptive genius of children deteriorate rapidly after they enter school, where they must sit for lengthy periods, often in chairs inappropriate for their size.

We know that it takes many months, even years, for monks to be trained to sit in stillness and to concentrate simply on their breath. It is not surprising that children, without any training whatsoever, gradually become dissociated to avoid the discomfort of their bodies.

This subject is actually relevant to the new technology because it is rarely addressed and also because it is a blind spot in our culture. A stroke patient or paralyzed patient being taught to use interface technology carries an underlying dissociation that will affect the experiment. In addition to designing these amazing new interfaces, it should be possible to improve the *internal* interface between the subject's body and mind. This is what the second person specialist knows how to do. It would be a waste not to implement this sophistication at the beginning of all these wonderful inventions.

4. Re-embodied cognition

We will now consider high performance psychophysical activity, a category which will include physical rehabilitation. (A physically disabled person and an athlete working to gain that winning second face similar proprioceptive challenges: in both cases, the refined skill takes place in micro movements.) The dedication necessary to relearn movements that for so long had been taken for granted is, in our view, a high performance discipline. The first step in the process is the re-association of mind and body. Dissociation is so common that it goes unnoticed until a crisis demands a change. Returning the mind to its moment by moment engagement with proprioception recalibrates the nervous system into psychophysical congruence.

A most obvious symptom of common dissociation is the lack of sensorial definition, the lack of a positive sense memory in particular. It is interesting that we all want to feel good, but when asked to be very precise in our definition of that desire, most are hard put to state sensory data clearly. When it comes to pain, we are often much more precise. No doubt it has been an evolutionary advantage to deeply embed memory of danger rather than pleasure, an adaptation for the purpose of long-term survival. In high performance activity, a break in congruence is often the most difficult obstacle, only surpassed by the dissociated internal negative critique.

What do these second person psychophysical experts do to re-embody the minds of people who are attempting to surpass Olympic records, sing at La Scala, or operate a wheelchair in a world made for walking?

They begin with basics. They analyze how the subject is using their autonomic nervous system, which either limits creativity in favour of immediate survival or opens the nervous system to aesthetic invention. Figure 3 diagrams the basic structure of the autonomic nervous system, with its two branches:

On the left, the sympathetic (fight or flight) system dilates pupils, inhibits salivation, accelerates respiration, accelerates heart beat, inhibits digestion, secretes nora-adrenaline, increases sweat, raises "goose bumps", relaxes bladder, and stimulates orgasm.

On the right, the parasympathetic (rest and digest) system constricts pupils, stimulates salivation, relaxes respiration, slows heart beat, stimulates digestion, decreases sweat, relaxes hair follicles, constricts bladder, and stimulates sexual arousal.

In order to re-embody the mind, it is necessary to know whether the sympathetic or parasympathetic system is dominant. Usually it is the sympathetic that is overactive. Of course, a subject in an experiment is going to be flooded with insecurities, doubts, fears of failure, fear of disappointing the researcher, etc. It is important to remember that most human beings in an unknown experience often become anxious. Their performance during an experiment therefore may not fairly represent their true capacities. Sympathetic dominance limits awareness of the present moment to that which is necessary for survival. Therefore, it eliminates aesthetics, subtleties, and the creative genius that is possible in the parasympathetic state. High performance requires an adaptable union of parasympathetic calm and excited precision of the sympathetic. If either one becomes too dominant, the *edge* of performance is lost.

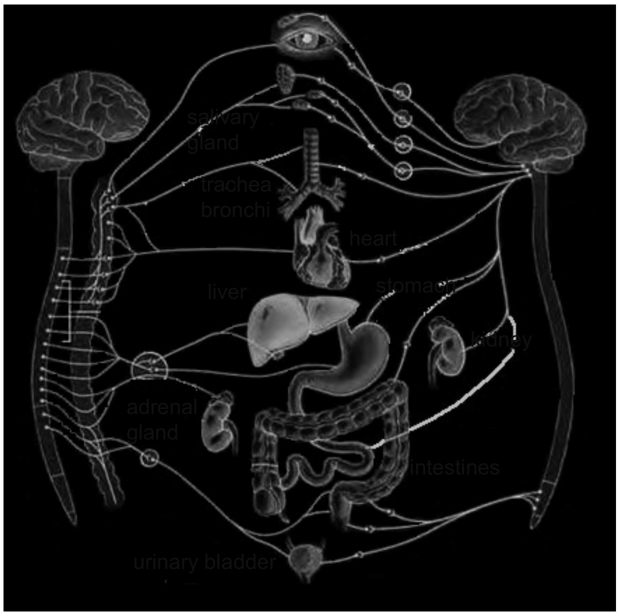


Figure 3. Sympathetic vs. Parasympathetic

In the case of an enactive interface laboratory subject, particularly someone like a quadriplegic who has already lost proprioceptive control of her existence, the stress induced by self-criticism always waits in the shadows. It anticipates the condescension of the technician, who may be so invested in technology that the enactive interface only receives lip service.

It was very encouraging to find a research focus on the laboratory subject's stress during the enactive interface XVR workshop. Guenter Edlinger (Guger Technologies) included stress feedback (via Galvanic skin response) in his experiments. The team of Roberta Carabalona and Paolo Castiglioni (Don Gnocchi Foundation, Italy) monitored the effect that a disruption in virtual technology, which they call *a break in presence,* has upon laboratory subjects. In the performing arts, a break in concentration is said to break the *intimacy of space.*

Psychophysical specialists often find that this sort of disruption is followed by an irrational, hallucinated fear and a shift into dissociated (third person) negative critique. Unfortunately, the resulting loss of performance can escalate in mere seconds. High performance training shows people how to return to a more parasympathetic state of awareness. In essence, psychophysical specialists have learned how to train the autonomic nervous system.

There is a link between Enactive Cognitive Science and psychophysical methodologies. In the book *The View From Within: First-person approaches to the study of consciousness* (edited by Varela), Carl Ginsberg mentions two methods in particular [14]. The first was developed by F.M. Alexander in the late 1800's, and he continued to refine his *Alexander Technique* in the early 1900's in England. [15]. One of his greatest supporters was none other than Sir Charles Sherrington, who received the Nobel Prize for his explanation of how it would be possible for the nerves, originating in the spinal cord, to communicate through neurons and synapses, and by so doing to organise the entire body in the maintenance of posture [16]. Sherrington also coined the term "proprioception."

The Alexander Technique is based upon a discovery of the importance of a gravity-sensitive (dynamic) positioning of the head in relation to the function of the spine. This discovery was verified by the work of R. Magnus [17] on *righting reflexes*, which he called *central control*, and later by Alain Berthoz [6]. Highly skilled teachers help students of the Alexander Technique retrain their natural head/neck coordination, stimulating a conscious awareness of this relationship as the student refines a new skill. High performance artists thrive with this kind of gentle but very precise psychophysical control. It makes their performance effortless and at the same time inspired.

The second method was introduced in the 1950's by Moshe Feldenkrais, a Russian, French-trained physicist and engineer who experimented with many different physical sports and studied the methods of those rare psychophysical specialists that existed in the first half of the 20th century. His genius was to combine sequences of infant proprioceptive dynamics, his knowledge of physical forces, and a profound understanding of the human nervous system. He invented 600 movement experiments that re-establish access to original proprioceptive intelligence. He discovered that by re-enacting movement patterns, one can re-learn or even learn for the first time steps in developmental coordination that might have been missed. These exercises are called *Awareness Through Movement.* [18].

The training received by teachers of these two methods is so far in advance of most physical therapies that they have a devoted following by students at the highest artistic and athletic performance levels. Both methods may be studied individually or in group lessons. The Alexander Technique is taught in every major theatre and music school in the world; the Feldenkrais method has also been integrated into dance and theatre training.

I have provided brief definitions of the two methodologies as background. I am not suggesting that specialists actually teach these methods during laboratory experiments. Rather, they are ideal providers of *second person feedback*. They can improvise on the spot, and mediate between all the interfaces in the loop of the experimental system. As an interdisciplinary group, they have begun to study the vocabulary and concepts of neuroscience and interface technology. They attend lectures and maintain dialogues with philosophers, technicians, researchers, cognitive scientists, and engineers. They educate each other in their conferences. Their primary *modus operandi* is non-intrusive intervention. I believe that their exceptional experience and dedication will be useful to scientists and engineers.

Just as the enactive interface engineers are at the leading edge of their disciplines, this growing group of psychophysical experts have a fundamental interest in eliminating the errors of disembodiment for the children of the future as well as assisting in the design of rehabilitative equipment for the disabled.

5. Conclusions

An embodied perspective on dynamic systems allows everything and everybody involved in the experimental system to be recognised. The rigor of this comprehensive view accepts the scientist's preference for a reductionist validation without sacrificing the significance of the human experience. We are now at a stage where interdisciplinarity must be the watchword of the future. Up till now, we have assumed that the third person position had attained its full identity. If we had spent the last few centuries exploring embodied objectivity, however, we would probably have discovered the state of *engaged detachment* without any loss of rigor—and with far less back pain. High performance specialists base their performance on a fluid perception of inner and outer experience.

During the 20th century, the highly respected American anthropologist Margaret Mead was a voice of reason that helped guide the American population through their post-war cultural upheaval. She was a founding member of the famous Macy Conferences, which gave birth to Cybernetics and opened the post-behaviourist path to Cognitive Science. If she were alive today, she would be fascinated but cautionary, reminding engineers of the responsibilities associated with transforming a culture. She would suggest that whether technology destabilises a culture is likely to depend on *how* it is introduced to the population: it could create more disembodiment, or it could improve human potential by stimulating psychophysical congruence.

From a sociological perspective, we have before us a graying population still struggling with computer fluency and mobile phones, while the next generations learn to open a computer program as easily as a refrigerator door. As children begin to relate to robotics in the same way that their grandparents may have related to television, we can expect an even more awkward generation gap to develop. Sensing the enormity of the change before us, a new transdisciplinary interest is awakening in both enactive cognitive science and psychophysical experts who feel a need to participate in whatever future interface engineers have the courage and the skill to create.

If Susan Greenfield, in her book *Tomorrow's People: How 21st Century Technology Is Changing the Way We Think and Feel* [19], is right about a future house computer that will monitor our life signals and operate as an external homeodynamic regulator of our well being, the second person expert is an essential catalyst. Such an interface cannot be accomplished with only a third person model.

The need is for technology that enhances congruent, embodied cognition. In this light, our responsibility is to consider a wider view of what could go wrong if truly embodied cognition is *not* included in future technology. Likewise, it falls to us to imagine what could go very right if enactive engineers sat down to plan a future with their neighbours, enactive cognitive scientists.

6. Figures

Figure 1. Francisco Varela: photograph courtesy of Amy Varela and Michel Bitbol.

Figure 2. Lower back pain, neck, migraines and face pain, in the past 3 months (2000). Sources: Center for Disease Control and Prevention, National Center for Health Statistics, United States, 2006, Fig. 30. Data from National Heath Interview Survey.

Figure 3. The autonomic nervous system: Psychology Image Bank, McGraw Hill.

7. References

[1] F.J. Varela, E. Thompson, and E. Rosch. *The Embodied Mind: Cognitive Science and Human Experience.* MIT Press, Cambridge, 1991.

[2] F.J. Varela and J. Shear. First-Person Methodologies: What, Why, How? *Journal of Consciousness Studies*, Imprint Academic, v. 6, no. 2/3, pp.1-14, 1999.

[3] E. Thompson. *Mind in Life: Biology, Phenomenology and the Sciences of Mind.* Harvard University Press, Cambridge, 2007.

[4] E. Thelen and L.B. Smith. *A Dynamic Systems Approach to the Development of Cognition and Action,* MIT Press, Cambridge, 1994.

[5] L. Eliot. *What's Going on in There? How the Brain and Mind Develop in the First Five Years of Life.* Bantam Books, New York, 1999.

[6] A. Berthoz. *The Brain's Sense of Movement,* Harvard Press, Cambridge, 2000.

[7] S. McCredie. *Balance: In Search of the Lost Sense,* Little Brown and Company, New York, 2007.

[8] H. Maturana and F.J. Varela. *Autopoiesis and Cognition: the Realization of the Living*, R.S. Cohen and M.W. Wartofsky (eds.). Boston Studies in the Philosophy of Science v. 42. , D. Reidel Publishing, Dordrecht, 1973, 1980.

[9] J. Cole. *Pride and a daily marathon*, Duckworth, London, 1991.

[10] D. Stern. *The Interpersonal World of the Infant: A View from Psychoanalysis and Development,* Basic Books, New York 1985 and 1998.

[11] T.C. Dalton and V.W. Bergenn (eds.). *Beyond Heredity and Environment: Myrtle McGraw and the Maturation Controversy*. Westview Press, Boulder, CO., 1995.

[12] G. Rizzolati. "Mirror Neuron System" Discourse given on 12 December 2006 at the *Réception des Associés étrangers élus en 2005*, *Institut de France*, *Academie des Sciences*.

[13] J. Schwartz. *Mind and Brain: Neuroplasticity and the Power of Mental Force*, Regen Books, New York, 2002.

[14] F.J. Varela and J. Shear (eds.). *The View From Within: First person approaches to the study of consciousness.* Imprint Academic, UK, 2000.

[15] J. Dewey. "Introduction" to FM Alexander's *Conscious Control of the Individual*, E. P. Dutton & Company, New York, and Centerline Press, 1985.

[16] C.S. Sherrington. *The Endeavour of Jean Fernel, with a List of the Editions of His Writings*, Cambridge University Press, Cambridge, 1946.

[17] R. Magnus. (1924) *Body posture = Körperstellung: experimental-physiological investigations of the reflexes involved in body posture, their cooperation and disturbances*. Edited, with a preface by A. Van Harreveld [New Delhi]: Amerind, Springfield, VA, 1987.

[18] M. Feldenkrais. *Awareness Through Movement: An Easy way to Improve Your Health, Vision, Imagination, and Personal Awareness.* Harper One, New York, 1991.

[19] S. Greenfield. *Tomorrow's People: How 21st Century Technology Is Changing the Way We Think and Feel*, Penguin Books, London, 2004.

Applications

A Multimodal and Enactive Interface for Aesthetic Shapes Evaluation

M. Bordegoni[*] U. Cugini[*]
(*)Dipartimento di Meccanica, Politecnico di Milano, Italy
E-mail: monica.bordegoni@polimi.it, umberto.cugini@polimi.it

Abstract

This paper presents a multimodal and enactive interface for the evaluation and modification of the shape of aesthetic products. The evaluation and modification applied to shapes is based on the manipulation of characteristic curves, which is a typical practice used by designers in the industrial design domain. The interface consists of a haptic strip that conforms to a curve that the designer wishes to feel and explore, or that can be locally or globally deformed in order to apply modifications to the corresponding curve, and consequently to the shape. The haptic strip is an innovative solution in the haptics domain, although has some limitations concerning the domain of curves that can be actually represented. In order to extend this domain and make users feel some curve features, for example curvature discontinuities, a multimodal interface has been implemented that adds sound as additional information channel.

1. Introduction

The evaluation of aesthetic aspects of new conceived shapes is an important phase in product development that may determine the ultimate success of the product. Current technologies allow designers to work in a digital context, where activities that traditionally have been carried out on physical mock-ups are today performed on virtual prototypes. The evaluation of aesthetic aspects and appearance of shapes is a task that is often performed by means of visualization tools that are today very sophisticated and performing. In fact, they support realistic rendering, as well as real-time and interactive rendering. The possibility to evaluate shapes only on the basis of the visualization of the product is actually limiting for the designers, who need to physically interact with the evolving shapes of the products they are designing [1], so as to check and evaluate properly all the aesthetic features of their products. In addition, if they do not like the shape and wish to modify it they would like to do it immediately and easily, and possibly within the same environment they use for evaluating the shape.

The possibility to physically touch and modify shapes can be provided by adding haptic interfaces to design tools. Most of the haptic interfaces oriented to the design domain that have been developed so far are of point-based type [2]. Actually, designers do not find intuitive using a stylus or a device alike for exploring a surface. In fact, they are used to freely pass their hands over shapes for detecting the quality of the surface. This capability they have in detecting surface defects using their hands is part of the designers' knowledge and skills.

So, in order to develop a new tool that can be effectively used by designers for the evaluation and modification of shapes a possibility is to equip a CAD (Computer Aided Design) tool -which offers good capabilities in terms of control of the mathematical quality of the geometric models and their realistic visual representations- with haptic interfaces that the designers find easy to use according to their skills [3].

The paper presents a system that we are developing within the context of the EU funded project SATIN that is based on a deformable physical strip that provides the possibility to touch and interact through a continuous curve/line that is interactively positioned on a virtual surface and that is co-located with a realistic 3D stereo view of the whole object being modelled [4]. The system can be used for evaluating and modifying shapes through the interaction with haptic representation of characteristic curves.

The idea of proposing a new CAD tool having a novel interface based on the haptic strip was generated by the observation that designers of aesthetic products create and manipulate shapes on the basis of curves: characteristic curves, aesthetic curves, etc. The concept of a curve as basis for shape generation and modification is not only used in digital models, but also in physical models (for example, the use of the red tape in car design).

The haptic interface is built on top of a CAD tool that typically provides an almost unlimited domain of shapes in terms of mathematical surfaces or solid models with a very high flexibility for modification and control of geometric characteristics, and in particular of continuities of various orders in the application domain of aesthetic objects.

The strip approach has an intrinsic limitation due to the fact that a continuous deformable strip is able to reproduce physically only a limited number of mathematical shapes. In fact, if the deformation of the continuous physical strip is realized by means of acting on a limited number of points that are positioned in space, then only span wise splines can be generated. Therefore, the issue to address is to find ways for extending the domain of representable shapes, or better of "perceived shapes".

The paper addresses this issue and discusses how multimodal interaction may convey geometrical characteristics of shapes that cannot be represented by current haptic technologies.

2. State of the Art on haptic interfaces and multimodal interaction in shape perception

This section describes the state of the art of haptic interfaces and the role of modalities during the perception of shapes and how multimodal interaction may support this task.

2.1 Haptic interfaces

Since the appearance of the first devices (around 15 years ago) [5] most of the research and development activities have concentrated on point based force-feedback devices with limited workspace and force feedback hardly matching real industrial working conditions, and much less on surface based tactile devices currently limited to proof of concepts of technological principles and very far away from prototypes allowing to demonstrate the possibility of addressing industrial applications. Anyway, the two aspects are kept separated mainly because of technical difficulties in considering both of them together.

Most of force feedback technologies developed so far is based on point-based contact (0D), and in certain cases, especially in the medical field, have reached a high level of efficiency for very vertical applications generally dedicated to training activities. The evolution of both research and industrial applications seems to be based on multi-point contact which is a way to improve and enlarge the field of application grounding and extending on the current experimented technology.

Some tentative devices providing the feeling of a contact with a surface has been more recently developed. They consist of small patches of sensors/actuators that can render very small surfaces (2D) and support limited input forces [6]. In addition, although the mechanical architecture of these systems is simple, it is difficult to imagine that the resolution and the dimension of the contact surface may increase at an application level also in the medium term of 5 to 10 years, due to the difficulty of extending the prototype up to a resolution equivalent to a visual display of the same size.

Our system (named SATIN system) aims to be a major advance in respect to the state of the art, since it would be the first device proposing a continuous contact along a line (1D).

In order to better clarify the novelty of this system it may be useful to position it with respect to current available technologies, prototypes and research works ongoing in the haptic field. Let us consider a 3D space characterized by the following three axes (as shown in Fig. 1) in which to map technologies and prototypes:

- Dimension of contact space on the virtual object: 0D point, 1D line, 2D surface.
- Dimension of contact space related to the user: single point, multiple points, continuous line, and continuous surface.
- Dimension of the interaction/working space: 1D, 2D, 3D.

We position in such space some of the most relevant haptic technologies: point-based devices like PHANToM [7], FCS-HapticMaster [8], and the Haption-Virtuose [9], multipoint based devices like the Haptex system [10], matrix-based [6], T'nD system [11-12]. Within the space it is positioned also the SATIN system based on a deformable strip and allowing a user to touch along a continuous line a virtual 3D object. The system provides a dimension of interaction that is not covered by other haptic technologies.

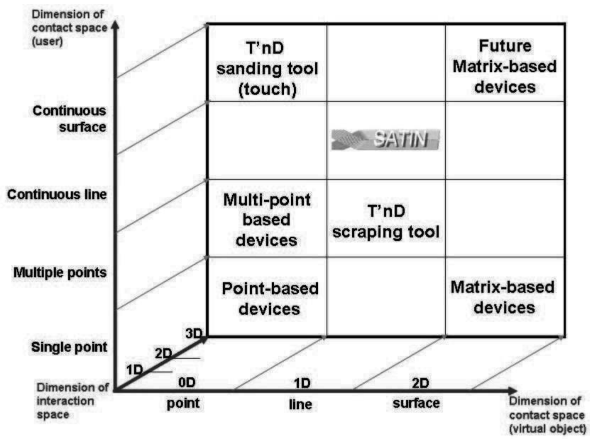


Figure 1. Overview of the haptic technology mapped according to rendering dimensions and application level.

2.2 Multimodal interaction

Modalities tend to perform in a complementary way to integrate information about an object. Lederman et al. [13] show that when assessing the physical nature of an object such as sandpaper, vision provides information about the compactness of the surface, whilst touch provides information about its roughness. Whilst the information from these modalities is different, they combine to provide an understanding about the physical nature of sandpaper. In this sense the modalities are *complementary*.

As well as being complementary, modalities are also integrated to provide the best possible information through a process termed ‘maximum likelihood estimation’. This is demonstrated by Ernst and Banks [14] who show that haptic and visual integration is statistically weighted dependant upon the quality of the signal received. In a situation where there is a lot of visual ‘noise’, haptic information tends to dominate and vice-versa. This may prove a useful facet where *high fidelity rendering of the prime modality is not possible, a secondary modality may be able to support object property perception*.

At times what is perceived visually can be ambiguous and is in need of resolution through alternate *modality confirmation.* This is particularly the case where depth cues are limited. For instance, Sekuler and Sekuler [15] demonstrate that when two discs are observed rolling towards each other, they will be perceived to either bounce off each other or pass behind each other at the intersection point. The scene in this sense is ambiguous, as both possibilities are visually plausible. However when sound is introduced and the balls are heard to collide the ambiguity of the scene is resolved.

Faconti et al. [16] have investigated visuo-haptic cross-modality interaction in presence of ambiguous visual stimuli that may conflict with haptic information. The experiments concluded that haptics is used to confirm the expectation of the subject’s mental model built from vision.

The presence of another modality can also provide greater *accuracy of judgement*. Guttman, Gilroy et al. [17] show the influence of sound in judging a visual stimulus. Observers were asked to view two visual sequences of light and dark bars and judge if they were the same. Accompanying the visual stimuli was a sound sequence; sometimes it was in time to the visual 'beats' whilst at other times it was out of sequence. When the ‘beats’ of vision and sound coincided pattern detection was almost perfect; when they were out of synchronisation results were at the level of chance.

Eventually, *redundancy* may be introduced where the same information can be conveyed by several different modalities. Illusions that are sometimes used as a means of distortion of the senses may play an important role in this context, where the intended information may be conveyed through the use of different modalities for “correcting” or “improving” messages that are partial or incomplete because of technical limitations of other channels.

3. Haptic strip

The SATIN system consists of an Augmented Reality environment where the user – a designer – can interact with 3D virtual objects. The user is able to see the object and to both explore and modify the shape of the object through the use of touch. The SATIN system consists of a haptic strip that mimics the tape placed by designers on physical mock-ups for evaluating characteristics and style lines (Fig. 2). In addition, a 3D visualization of the shape is super-imposed onto the physical device by means of a stereoscopic display system [18].

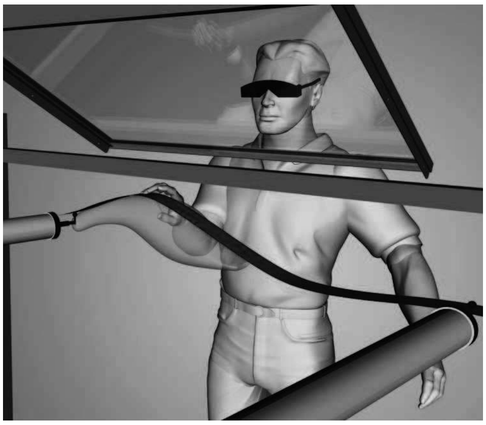


Figure 2. Conceptual image of the haptic strip for shapes evaluation and modification.

The haptic interface is a major advance in respect to the state of the art, since it is the first device proposing a continuous contact along a line. The main problem we have faced in the design and development of the SATIN haptic interface is providing a tangible experience satisfying two aspects:

- having a long continuous deformable haptic surface to be touched (to perceive the global shape);
- having the possibility to communicate (make the user perceive) local and very fine geometric characteristics as curvature discontinuities, which are considered as surface defects in product design.

The domain of shapes (surfaces and curves) that we can represent in surface modelling systems is much larger than the domain of shapes that we can produce physically by deforming a continuous body such as the strip we are developing. Preliminary studies about haptic technological solutions have clearly shown that we are not able to obtaining a one-to-one representation of mathematical shapes through a physical strip. In order to properly design the physical strip and define its number of degrees of freedom, we had the necessity to study human perception capabilities related to curve feature perception and discrimination. The perception of the quality of a curve is a subjective process that is based on the skill related to the sense of touch of the user, and in addition involves the perceptual sphere of humans.

For what concerns the physical evaluation of the quality and characteristics of a curve, we performed a set of experiments in order to quantify the real perceptual accuracy and sensitivity of humans; in other words, we wanted to understand what is the human perceptual threshold of curvature discontinuity of curves and surfaces [4]. It was evident from the experiments that there is not a major human capability in discriminating curves properties and perceiving differences. Therefore, we can assume that it is not necessary to reproduce the physical shape of curves with high precision. This conclusion is being used for designing the SATIN haptic strip.

The haptic interface of the SATIN system consists of a haptic strip made of a plastic tape actuated by 9 servo drives that is connected to a 6DOF platform consisting of two FCS-HapticMaster systems (FCS-HapticMaster) operating in a parallel configuration. The basic concept is to use as main user interface a force sensitive tangible strip, suspended in space in the position of a section of a simulated virtual object. The strip can actively shape itself, and place itself in the appropriate position and orientation in the workspace. The psychophysical requirement to satisfy is to be able to explore and modify this part of a virtual object by touch. As a shape exploration tool, the system represents one of the very few attempts at a full, whole hand, encountered shape display, but with the limitation that it represents only a planar cut through the surface.

One of the key features of the SATIN concept is that the tangible strip is at the same time an output device, and an input device. As an output device, the strip is an exploration device for the human hand and fingers to touch. As an input device, the strip behaves as a physical item which can be shaped by hand like a physical bending spline. The haptic strip is implemented by means of continuous physical spline that is actuated into the desired shapes by nine equidistant relative actuators along its length. Fig. 3 shows a conceptual image of the strip mechanism using nine servo drives.

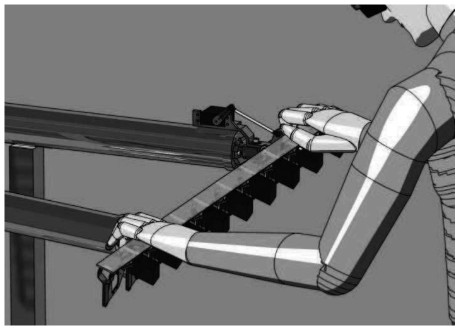


Figure 3. Detail of the strip mechanism with 9 servo drives.

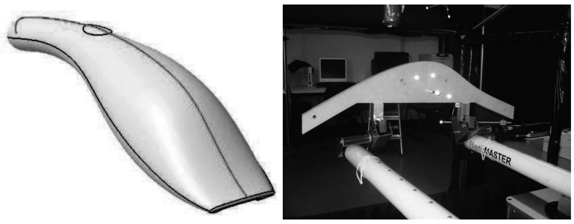
The visualization solution consists of a DLP projector that projects the light onto and through a set of mirrors and screens forming a 3D image of the virtual object on top of the haptic strip [18]. The projector is located above and on the back in respect to the haptic system, so that it does not occlude the image projected by the projector. The display system is designed in a way that its components do not interfere with the haptic workspace. So, the user can freely move his hands within the interaction space, and is able to interact with the system by grabbing and manipulating the physical interface that is positioned under the mirrors (Fig. 6).

4. Extending the domain of perceived shapes

The domain of shapes (surfaces and curves) that we can represent in surface modelling systems is much larger than the domain of shapes that we can produce physically by deforming a continuous body such as the strip we have developed. Because deforming a continuous strip of material is not possible (or at least we were not able to find a way) in order to reproduce a punctual discontinuity the basic idea of the project has been to generate an haptic illusion using sound by mixing force cues with sound cues to make people feel shapes that differ from the actual shape of the object.

An important and innovative element of the SATIN system lies in the use of sound as a means to convey information and feedback about the virtual object and the user interaction. More specifically, the use of sound allows the designer to explore geometric properties of the object that are not detectable by touch or sight. The SATIN system is based on an enactive interface that is capable of conveying information about shapes to users and understanding users' actions on shapes in order to provide an adequate response in visual, haptic and sound perceptual terms. SATIN exploits skills of expert designers who work normally with their hands.

The concept of mixing haptic and sound has been validated through a mock-up of an audio-haptic interface used for performing some tests with users. The audio-haptic interface setup for testing consists of a physical template, which simulates the future haptic strip. The template consists of a profile generated by the symmetry plane of a vacuum cleaner (Fig. 4).



a) b)

Figure 4. a) Vacuum cleaner and profile selected. b) Physical template of vacuum cleaner section.

The strip is covered with a sensorised tape computing the position of the finger in real-time. Geometric models compute local characteristics to be mapped to sound parameters in order to provide an audio feedback to the user. Currently, the sound application is implemented for playing the curvature of the surface. Three distinctive aspects of these data are explicitly considered:

- the absolute value of the curvature, indicating how strongly curved the surface is;
- the sign of the curvature, indicating whether the surface is concave or convex;
- potential discontinuities in the curvature.

These parameters are mapped to different acoustical parameters used to control or modify a sound source. In the initial application, a pure tone or sine wave is used as the source.

The absolute value of the curvature data is mapped to the frequency of the pure tone. The sign of the data is mapped to the stereo panning of the sound output in the following way: if the data are negative, the output is weighted to the left channel, otherwise to the right channel. To signify discontinuities in the curvature data, two different modulations are applied to the sound.

When the user moves his finger across a discontinuity position of the surface, the amplitude of the audio output is rapidly increased and then decreased, resulting in a click-like sound played in synchrony with the finger's movement across the discontinuity. This sound is easily distinguished from the pure tones associated with "ordinary" curvature values and gives the user initial feedback about the presence and the approximate position of the discontinuity.

A series of tests have been performed using the audio-haptic system (Fig 5). The testers have been asked to perform the following tests:

1. Move one finger over the physical template and evaluate the correspondence of sound variation with the variation of physical curve geometric characteristics;
2. Move one finger over the physical template and say if and when discontinuities are perceived.

The testing results reported that in general sound was viewed positively, and that it showed discontinuities well. Users generally felt comfortable using the system, users were able to make precise movements and performed the task well. Participants also reported understanding the meaning of the sounds. The users did not find the system difficult to learn, no lag in system response was perceived and the response of the system met user expectations.

Figure 5. Testing session of the audio-haptic mock-up.

Due to the positive results obtained in the users' testing, the audio-haptic interface will be included within the SATIN system that is shown in Fig. 6.

5. Conclusions

The paper presented a system developed within the context of the European project SATIN – Sound And Tangible Interfaces for Novel product design (http://www.satin-project.eu) whose aim is allowing designers of aesthetic products to use virtual prototyping techniques to shorten time-to-market, adding to current virtual prototyping tools the possibility to touch and modify virtual objects by hands and exploiting their acquired skills. Designers can perceive, evaluate and modify an object shape along a curve by interacting with a haptic strip, looking at the stereoscopic visualization of the object and also hearing as sound the relevant geometric characteristics of the shape under construction.

The haptic interface offers the possibility on one side of evaluating early in the design process components that do not exist already, and on the other side of better feeling components trough physical interaction. This practice can be effectively used for rapid design evaluation and review of new products. The major benefit of the system would be the reduction of the time required for the production of physical prototypes, and therefore the reduction of the total time-to-market, without affecting the quality of the final product.

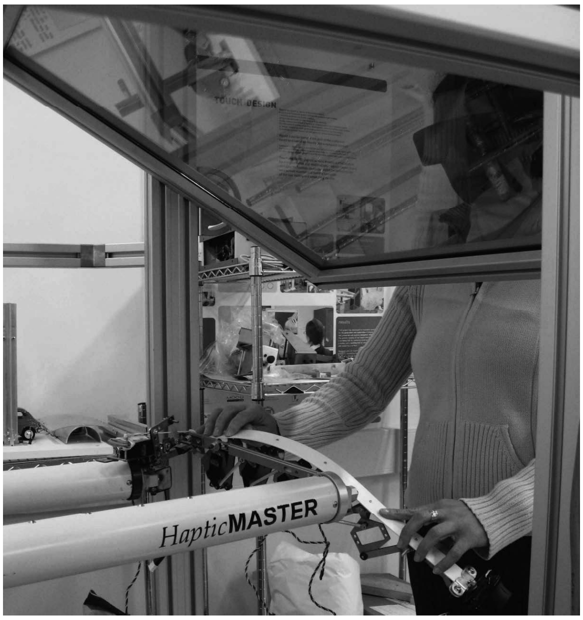


Figure 6. Prototype of the SATIN system.

6. Acknowledgment

The research work presented in this paper has been partially supported by the European Commission under the FP6-IST-5-054525 project SATIN – Sound And Tangible Interfaces for Novel product design (http://www.satin-project.eu). The authors would like to thank all the project partners for their contributions to the research.

References

[1] Y. Yamada, *Clay Modeling: techniques forgiving three-dimensional form to idea*, San'ei Shobo Publishing Co., 1977.

[2] G. Burdea, Force and Touch Feedback for Virtual Reality, New York: John Wiley & Sons, 1996.

[3] U. Cugini, M. Bordegoni, Exploitation of manual skill for the generation of aesthetic shapes: a typical enactive approach, *Proc. of ENACTIVE/07, 4th International Conference on Enactive Interface*s, Grenoble, France, November 19th-22nd, 2007.

[4] M. Bordegoni M., U. Cugini U., Issues on aesthetic shapes perception for the development of new design tools, *Proc. of the ASME - IDETC/CIE Conference*, Brooklyn, New York, USA, 2008

[5] T.H. Massie and J.K. Salisbury , The PHANTOM Haptic Interface: A Device for Probing Virtual Objects, *Proc. of the ASME Winter Annual Meeting, Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, Chicago, IL, 1994.

[6] H. Iwata, H. Yano, F. Nakaizumi, R. Kawamura, Project FEELEX: Adding Haptic Surface to Graphics, *Proc. of SIGGRAPH 2001*, Los Angeles, CA, 2001

[7] PHANToM device, SenSable Technologies Inc., http://www.sensable.com

[8] FCS-HapticMaster, MOOG-FCS, http://www.moog-fcs.com/robotics

[9] Virtuose, Haption, http://www.haption.com

[10] Haptex system, http://www.percro.org

[11] M. Bordegoni, M., Cugini, U., Haptic modeling in the conceptual phases of product design, in *Virtual Reality Journal, Special Issue on Haptic Interfaces and Applications*, Springer Computer Graphics Format, 9(1).

[12] M. Bordegoni and U. Cugini, A conceptual design tool based on innovative haptic interfaces, *Proc. of DETC'06 ASME 2006 Conference*, Philadelphia, Pennsylvania, USA, 2006.

[13] S.J. Lederman , G. Thorne et al. Perception of texture by vision and touch: Multidimensionality and intersensory integration. in *Journal of Experimental Psychology: Human Perception and Performance*, 12(2): 169-180, 1986.

[14] M.O. Ernst, M.S. Banks, *Humans integrate visual and haptic information in a statistically optimal fashion*. Nature 415(6870): 429-433, 2002.

[15] A.B. Sekuler, R. Sekuler, Collisions between moving visual targes: What controls alternative ways of seeing an ambiguous display? In *Perception*, 28: 415-432, 1999.

[16] G. Faconti, M. Massink, M. Bordegoni, F. De Angelis, S. Booth, Haptic Cues for Image Disambiguation, in in *Eurographics 2000*, Volume 19, Number 3, 2000.

[17] S.E. Guttman, L.A. Gilroy et al., Hearing what eh eyes see: Auditory encoding of visual temporal sequences, in *Psychological Science*, 16: 228-235, 2005.

[18] M. Bordegoni, U. Cugini, M. Covarrubias, Design of a visualization system integrated with haptic interfaces, *in Proc. of the TMCE 2008*, Izmir, Turkey, April 21–25, 2008, Edited by I. Horváth and Z. Rusák, 2008.

The voice painter

Amalia de Götzen Riccardo Marogna Federico Avanzini
Sound and Music Computing group
Dep. of Information Engineering, University of Padova, Italy
E-mail: {degotzen, marognar, avanzini}@dei.unipd.it

Abstract

Very often when looking at a painting or touching a sculpture it is possible to experience the meaning of being a perceiver as enactor of perceptual content. The piece of art that we try to explore forces us to move around the object in order to discover new meanings and sensations. We need to interact with the artistic object in order to completely understand it. Is it possible to think about sound and music in this way? Is the auditory and musical experience prone to such investigation? This paper tries to describe an enactive system that, by merging the concepts of autographic and allographic arts, transforms the spectator of a multimodal performance into the performer–perceiver–enactor. The voice painter, an instrument to paint with our voice movement in a closed loop interaction, offers a new artistic metaphor as well as a potentially useful tool for speech therapy programs.

1. Introduction

The topic of this work is inherently multidisciplinary: in order to discover and emphasize the enactive approach in artistic productions (dance, music, painting, sculpting etc.) by means of new technologies it is necessary to bridge the gap between technology and art, taking into consideration suggestions and needs from artists and constraints and possibilities from technicians. Arts and enaction can be considered strictly related even in classical works, where the technology is simply a brush and some colors, or the material of a sculpture and the tools used by the artist to build his piece of work. Two well known examples are the works of Close in painting [8] and Serra in sculpting [18].

However, new technologies for artistic expression that contemplate multimodal interaction give to the artist new tools and new ways to think about their work, involving the final users in an enactive interaction while experiencing a specific work of art. The idea of the *voice painter* device presented in this paper is to have simple – non-intrusive – technology that stimulates the user-perceiver-enactor to use her/his body and in particular his voice in order to create a painting, and discover the relation between her/his actions and the signs that she/he can produce. We think that this kind of approach is genuinely enactive and that it suggests a "third path" between allographic and autographic arts.

The paper is organized into three main sections. Section 2 summarizes the state of the art and main views about how enactive experiences inform artistic representations. In Sec. 3 the voice painter system is presented and described in its technical details while Sec. 4 is devoted to discussing applications of the system.

2. Autographic and allographic arts

One of the main categorizations between different forms of arts is the one introduced by Goodman [11] which defines 'autographic' and 'allographic' arts:

> the former cannot be noted and do not contemplate performance, while the latter can be translated into conventional notation, and the resulting 'score' can be performed with a certain freedom of variation.

Painting and music are the two artistic expressions that are generally used to exemplify this distinction. It is difficult to determine the rules that generated a given painting, there is no notation that can help someone else to produce an exact replica of an original piece of art: it is even possible to define every copy a 'forgery'. In music the point of view is totally different: every copy/performance of a piece is a possible interpretation. Notation allows many different musicians to play a given piece of music: the 'discrete' musical signals are first notated by the composer and then interpreted by the musicians. One can say that while autographic arts are one-stage arts, allographic arts are two-stage arts. The distance between these two forms of art can be dramatically reduced in modern performances, where, for instance, a painting can be seen as the result of a live performance: a dancer that paints with her/his body or a musician that controls some multimodal device that produces a video output while playing.

2.1 Enaction in Arts

One of the main achievement of the ENACTIVE project has been a deep and fruitful reflection on the role of enaction in the artistic creation process [3]. The topic is particularly hard to address since it links together abstract concepts which are difficult to define (enaction, creation): the Enactive/07 Conference in Grenoble has collected several contributions that can be analyzed in order draw a sort of "red line" that goes across different artistic expression with the common intention of exploring the enactive creative process.

The enactive theory of perception states that it is not possible to disassociate perception and action schematically, since every kind of perception is intrinsically active and thoughtful. In this view, experience is something that an animal *enacts* as it explores its environment [24, 19]. In this view, the subject of mental states is the *embodied* animal, situated in the environment.

Enactive knowledge can be acquired also when discovering a painting or a sculpture if the perceiver is immersed in this action-perception loop. The typical example of an enactive art is music: a violin player needs to feel and to hear the sound in order to adjust the performance. Following this perspective many enactive artistic applications created with the support of technology explore virtual instruments through different kinds of gestures or postures. At the same time these applications have to consider the specific feedback received e.g. when exploring a surface or when using a bow on a string [5]: we perceive through our hands and fingers a specific haptic sensation that stimulates the user/player to react in order to understand. Virtual musical instruments are then augmented with haptic devices that can render the surfaces and the forces involved while playing a real instrument.

2.2 Painting with voice

The use of the voice as enactive instrument is rather unexplored, particularly so in the context of artistic applications. Voice is a universal human-to-human communication means and is used to convey also non-verbal, paralinguistic elements including emotion, prosody, stress. Moreover voice and speech are always accompained by other non-verbal communication channels, such as facial expression, gesture, and body language, forming a single system of communication [17].

These observations provide the motivation for the development of an interface that uses vocal expression as a tool for creating visual signs. The central idea is to exploit the most relevant features of vocal expression and map them into graphic features, thus creating a simple and versatile instrument that can be used by an experienced performer as well as by a naif user. Similar ideas have been recently explored in [13] and [16]. In [13] the aim is the development of a drawing system for users with motor impairments, therefore voice is used as a controller (a "vocal joystick") which assigns different non-verbal sounds to different directions or actions. Due to the specific application field, the user voice is the only allowed input. Since the present work is focused on a different application field, it does not have this constraint and full body movement is used as a second input. Moreover, the mapping presented in [13] does not follow a *phonesthetic* [15] approach similar to the one presented here (except for loudness mapping).

The work presented in [16] shares more similarities with our main concept. In the concert performance *"Messa di Voce"*, the sounds produced by two vocalists are augmented in real-time by interactive visualization software. The mapping approach is similar to the one presented here, although in some scenarios more *iconic* mappings are used instead. The main difference lies in the localization/tracking system: while in [16] this is based on cameras and computer vision techniques, we introduce a microphone array-based localization system: in this way tracking is based exclusively on the incoming voice direction, and a single audio input system provides all the information for the mapping procedure.

In our scenario, the player will be able to paint on a black screen using her/his voice. The mouth can be considered the brush: in order to draw on the entire screen surface the user will be forced to move, therefore involving the whole body rather than just the voice as an input instrument. The action-perception closed loop is then recreated with the help of a system that will be presented in the next section.

3. System description

The system integrates an 8-microphone array and TDOA (time delay of arrival) estimation technique for localization and tracking of the user position (see Fig. 1). The user is supposed to move in the active area at a given distance from a screen. The graphic rendering is projected onto the screen in order to satisfy a full correspondence between voice source position and rendering position. In this way, the user is supposed to have

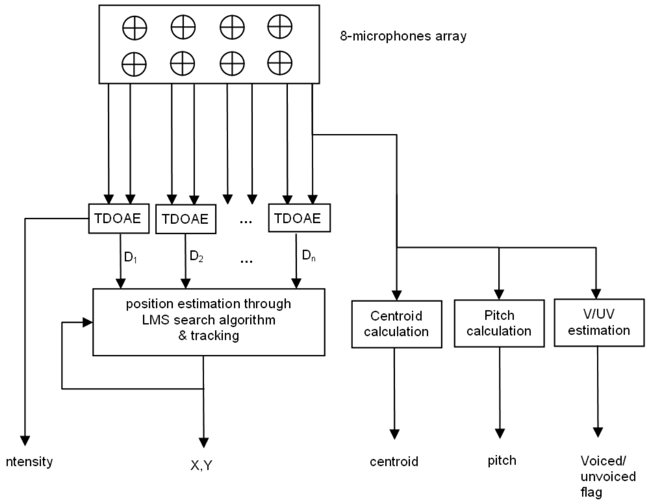


Figure 1: Block scheme of the system.

a more natural interaction with the virtual canvas, since the visual feedback is temporally and spatially correlated with the voice event.

3.1 Voice tracking system

The audio system performs real-time user localization and tracking through a 2-stage algorithm. First, TDOA estimation is performed for each pair of microphones. Then, the obtained vector of delays is processed in a LMS (*Least Mean Square*) algorithm in order to extract the estimated position.

The TDOA estimation stage implements an ATF (Acoustical Transfer Function) ratio estimation algorithm [10]. In a previous work [6] this method was recognize to achieve the best performances compared to other algorithms. The main idea behind the algorithm is to exploit some peculiar voice features in order to extract the signal of interest from the background noise and track it in a robust way. We made the assumption that background noise can be impulsive (e.g., a closing door) or stationary (fans, air conditioning systems,...). Noise events are therefore stationary or very short in time, while human voice is quasi-stationary (it can be considered stationary in a short-time frame of $20 - 30$ ms, while it changes its statistics from frame to frame). The algorithm exploits this assumption in order to estimate the TDOA for a given pair of received signals. Therefore, a vector of delays (VOD) is obtained.

In the second stage, the estimated VOD is compared to the elements of a pre–computed VOD matrix. This matrix is obtained through a discretization of the vertical plane of interest in the active area, and the subsequent calculation of the VOD for each position. Searching for the best fitting VOD in the search matrix with LMS criterion results in an estimated (x, y) position in the vertical plane of interest.

Although this stage achieves a robust position estimation (since it tends to ignore erratic delays and searches for a coherent figure), its main drawback is the poor performance due to the time-consuming search through the entire VOD matrix. This limitation can be improved, however, using a tracking technique. Since the voice source is supposed to move slowly in the active area with respect to the localization algorithm, position estimates that are close in time are supposed to be close in space too. Therefore, the second stage of the algorithm is improved by restricting the search to neighboring positions in an area surrounding the previous estimation. The exhaustive matrix search is then performed only at voice attacks. The described localization technique was implemented as an external module written in C for the Pure Data platform [20]. In this way, the algorithms have been highly optimized for real-time processing.

3.2 Feature extraction

Human voice can be characterized by several features, and human beings are able to control vocal emission in order to modify most of these features. We identified some prominent features which were relatively straightforward to be extracted with real-time algorithms. They were:

- intensity, computed as the RMS value of the squared voice pressure signal;
- spectral centroid, i.e. the center of gravity of the spectral magnitude computed over an audio frame;
- a voiced/unvoiced flag, depending on whether the utterance is associated to pseudo-periodic vocal fold vibrations or not;
- pitch, i.e. the subjective attribute of sound height.

Although the first two features are defined as usual and easy to extract, the other two are not so trivially estimated and therefore they are briefly described here. The voicing flag indicates the presence of a voiced signal, i.e. a signal containing periodicities due to vocal fold vibrations. This kind of signals can be detected with various approaches. We implemented a technique which combines zero-crossing detection and cepstrum extraction [1]. Pitch estimation is facilitated in this case by the relatively simple harmonic structure of voiced utterances, so that the problem reduces to estimation of the

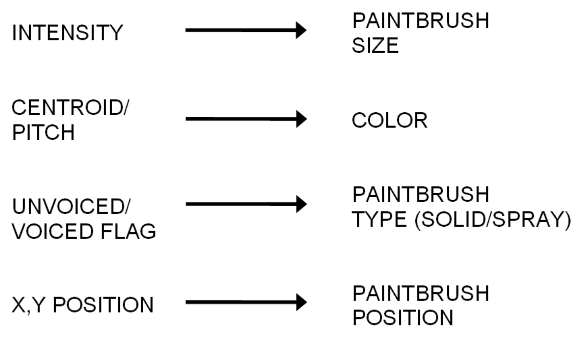


Figure 2: Mapping of voice features into graphic features.

fundamental frequency. This is estimated using an algorithm that extracts and matches harmonic spectral components on successive frames of the vocal signal [21].

3.3 Graphic rendering and feature mapping

Real time graphic rendering was performed using GEM [26], an OpenGL library designed to be integrated on Pure Data. The main idea was to construct a mapping of vocal features into well-recognizable graphic features (size, color, geometry). The choice of the mapping is crucial in designing the interface and dramatically influences the ability to control the digital instrument. Mapping strategies are the subject of several works in HCI literature and particularly for the design of digital musical instruments [14, 25].

As a starting point, the initial mapping was organized as in Fig. 2. It can be noticed that, although some features have a somehow immediate and intuitive mapping, for some other features such as pitch or voicing the corresponding graphic effect is quite arbitrary. For example, establishing a correspondence between sound frequency content and light frequency may not necessarily seem natural to the user, even though it may seem logical on numerical grounds.

The resulting visual effect is a sort of *abstract sketch* which can contain well defined geometric elements produced by short voiced segments, and/or particle-like signs due to unvoiced segments. The intensity and pitch/centroid information results in different color gradients and sizes. A snapshot of the resulting graphic rendering is shown in Fig.3.

Different and possibly more intuitive color mappings could be obtained e.g. in a RGB color field, by associating lower pitches to "hot" colors and higher pitches to "cold" ones. Similarly, HSV or HSL scales could be used to fit the pitch class scale. There is hardly any literature about applications that use this particular mapping and compare different possible strategies, therefore a simple one-to-one mapping was used in the first implementation.

An evaluation of this mapping can be based only on several subjective tests which are planned in the near future. Users will be asked to paint with their voice for 10 minutes to experience the interface. At the end of this preliminary training session they will be asked to perform several tasks using different mapping strategies. All tasks will concern the reaching of specific colors using the voice painter. Finally, the users will be asked to assess the quality of the mapping strategy at hand. The data will be then analyzed: timings, precision and paths will be collected and correlated to provide a quantitative assessment of each mapping strategy.

4. Discussion

A first informal test of the system was conducted with several users during various demo sessions (including the Enactive/07 Conference [3]). Users were allowed to interact with the virtual canvas without any hint about the graphic mapping and without any specific task (the only suggestion was: "use your voice to paint"). The goal of the test was to evaluate:

1. the way the user approaches the canvas and explores its voice features;
2. how many vocal features the user is able to identify in the mapping;
3. whether she/he is able to control an identified feature in order to obtain a desired graphic effect.

At the end of the test the user was asked to judge if the experience was natural for her/him, and in which cases.

This preliminary test showed that most of the features were identified by the users (except for the mapping pitch/centroid→color, which was not clearly recognized in several cases). It also confirmed that some mapping assumptions were natural for most of the users (e.g. intensity→size, mouth position→paintbrush position, voicing type→paintbrush type).

4.1 Robust speaker localization/tracking

Current developments are targeted at improving the robustness of the localization and tracking subsystem, especially in noisy, real-life scenarios.

The TDOA estimation and LMS algorithms described in Sec. 3.1 only consider impulsive or stationary noise background, and the performance is degraded

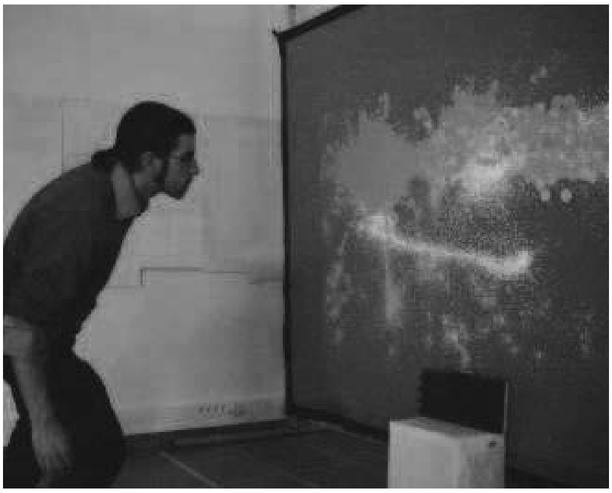


Figure 3: Performing with the voice painter.

in the presence of low SNR values especially when the noise is non-stationary. In order to improve robustness the first improvement in the tracking procedure is the inclusion of a more general voice activity detection (VAD) algorithm, that employs higher order statistics in order to allow detection of presence/absence of human speech in any region of audio [23].

A bimodal system is also being developed, in which localization and tracking is performed not only from the voice signal, but also using video. A preliminary implementation has been realized using the EyesWeb software platform for gesture analysis [7].

One more challenging issue concerns the problem of simultaneous localization of multiple acoustic sources, which would open up new possibilities for applications in cooperative scenarios. This will require the use of advanced statistical methods such as particle filtering [2].

4.2 Applications in speech therapy

Besides applications in performing arts and entertainment, the proposed system has the potential to be employed as an assistive technology in the field of speech therapy. HCI techniques are increasingly used as a method to teach and reinforce vocalization and speech skills in various contexts. It is generally acknowledged that interfaces with visual and multimodal biofeedback can influence the communication of individuals and can help facilitate in particular the speech and vocalization education process for children with communication skill deficits, both by motivating and rewarding vocalization and by providing information about the acoustic properties of vocalizations [4, 9, 12, 22].

Areas of application include *speech disorders* due to physical impairments or problems in motor planning and coordination (dysarthria, dyspraxia). Previous works have shown that computer-based speech training systems can help to improve articulation [4], and to attain correct production of specific sounds [22]. A second potential area of application concerns *speech delay* problems (i.e. situations in which speech development follows the usual patterns but at a slower rate than normal). In this case previous works have demonstrated the utility of systems that, through real-time analysis of vocalizations and appropriate biofeedback, reinforce of the production of syllabic utterances associated with later language and cognitive development [9]. A third potential area of application concerns problems associated to communication, social functioning, and expression, such as autistic spectrum disorder (ASD). Applied behavior analysis techniques are typically used, in which target behavior is rewarded e.g. with toys and the rewards are then gradually removed over time. This approach has been recently adopted in the design of computer-assisted systems that encourage playful behavior via technology, with the additional advantage that technology and computers reduce the apprehension caused by human-to-human interaction [12].

5. Conclusions

This paper presents an interface that allows explorations in different directions. The voice painter was born with artistic applications in mind: to create an instrument that could mediate between allographic and autographic arts, allowing sophisticated performances based on a musical notation (used then to paint) as well as improvisations or simple entertainment. The instrument has been first demonstrated at the Enactive/07 Conference and has spurred an active interest from the public and in particular among visual artists. Several different applicative scenarios have provided ideas for further development of the instrument: besides artistic applications, those related to speech theraphy and communication disorders are the most promising ones.

6. Acknowledgments

We would like to thank Barbara Mazzarino and Giovanna Varni for their useful suggestions and their help in developing the first implementation of the system, presented at the Enactive/07 Conference in Grenoble.

References

[1] S. Ahmadi and A. S. Spanias. Cepstrum-based pitch detection using a new statistical v/uv classification algorithm. *IEEE Trans. on Speech and Audio Processing, vol. 7, no. 3*, May 1999.

[2] F. Antonacci, D. Riva, M. Tagliasacchi, and A. Sarti. Efficient localization and tracking of two acoustic sources using particle filters with swarm intelligence. In *Proc. EURASIP European Sig. Process. Conf.*, Poznan, 2007.

[3] Association ACROE, editor. *Proc. 4th Int. Conf. on Enactive Interfaces (ENACTIVE/07…Enaction_in_Arts)*, Grenoble, 2007.

[4] O. Bälter, O. Engwall, A.-M. Öster, and H. Kjellström. Wizard-of-oz test of artur: a computer-based speech training system with articulation correction. In *Proc. 7th Int. ACM SIGACCESS Conf. on Computers and accessibility (ASSETS'05)*, pages 36–43, Baltimore, 2005.

[5] C. Cadoz. Musical creation process and digital technology. the supra–instrumental gesture. In *4th International Conference on Enactive Interfaces (Enactive'07)*, Grenoble, France, November 2007.

[6] G. Calvagno, C. Trestino, M. Romanin, and R. Marogna. A dsp implementation of a real-time speaker localization system using atf-ratio time delay estimation. *EDERS*, pages 1–10, Sept. 2006.

[7] A. Camurri, B. Mazzarino, and G. Volpe. Analysis of expressive gesture: The EyesWeb expressive gesture processing library. In A. Camurri and G. Volpe, editors, *Gesture-based Communication in Human-Computer Interaction*. LNAI 2915, Springer Verlag, 2004.

[8] C. Close, R. Storr, K. Varnedoe, D. Wye, and G. D. Lowry. *Chuck Close*. The Museum of Modern Art, New York, 2002.

[9] H. Fell, C. Cress, J. MacAuslan, and L. Ferrier. visiBabble for reinforcement of early vocalization. In *Proc. 6th Int. ACM SIGACCESS Conf. on Computers and accessibility (ASSETS'04)*, pages 161–168, Atlanta, 2004.

[10] S. Gannot, D. Burshtein, and E. Weinstein. Signal enhancement using beamforming and nonstationarity with application to speech. *IEEE Trans. Signal Processing, 49 (8)*, 2001.

[11] N. Goodman. *Languages of Art*. Hackett, 1985.

[12] J. Hailpern, K. Karahalios, J. Halle, L. DeThorne, and M.-K. Coletto. Visualizations: speech, language & autistic spectrum disorder. In *Proc. ACM Computer-Human Interaction Conf. (CHI'08)*, pages 3591–3596, Firenze, 2008.

[13] S. Harada, J. O. Wobbrock, and J. A. Landay. Voicedraw: a hands-free voice-driven drawing application for people with motor impairments. In *Assets '07: Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, pages 27–34, New York, NY, USA, 2007. ACM.

[14] A. Hunt, M. M. Wanderley, and M. Paradis. The importance of parameter mapping in electronic instrument design. In *NIME–02: Proceedings of the 2002 Conference on New Instruments for Musical Expression*, pages 149–154, Dublin, Ireland, 2002. Media Lab Europe.

[15] W. Khöler. *Gestalt Psychology*. Liveright Publishing Corporation, New York, 1927.

[16] G. Levin and Z. Lieberman. In-situ speech visualization in real-time interactive installation and performance. In *Proceedings of The 3rd International Symposium on Non-Photorealistic Animation and Rendering*, Annecy, France, June 2004.

[17] D. McNeill. *Gesture and Thought*. University of Chicago Press, Chicago, 2005.

[18] K. McShine, L. Cooke, J. Rajchman, B. Buchloh, and R. Serra. *Richard Serra Sculpture: Forty Years*. The Museum of Modern Art, New York, 2007.

[19] A. Noe. *Action in perception*. MIT press, Cambridge, Mass., 2005.

[20] M. Puckette. Max at seventeen. *Computer Music J.*, 26(4):31–43, 2002.

[21] M. Puckette and T. Apel. Real-time audio analysis tools for pd and msp. In *Proc Int. Computer Music Conf.*, pages 109–112, San Francisco, 1998.

[22] L. I. Shuster, D. M. Ruscello, and A. R. Toth. The use of visual feedback to elicit correct /r/. *Am. J. Speech-Language Pathology*, 4:37–44, 1995.

[23] S. G. Tanyer and H. Ozer. Voice activity detection in nonstationary noise. *IEEE Trans. Speech Audio Process.*, 8(4):478–482, 2000.

[24] F. Varela, E. Thompson, and E. Rosch. *The Embodied Mind*. MIT Press, Cambridge, MA, 1991.

[25] M. M. Wanderley and N. Orio. Evaluation of input devices for musical expression: Borrowing tools from HCI. *Computer Music J.*, 26(3):62–76, 2002.

[26] J. M. Zmölnig. Gem, http://gem.iem.at.

Representation of Samba dance gestures, using a multi-modal analysis approach

Luiz Naveda* and Marc Leman*
(*) IPEM – Dept. of Musicology, Ghent University, Belgium
E-mail: luiznaveda@gmail.com, Marc.Leman@UGent.be

Abstract

In this paper we propose an approach for the representation of dance gestures in Samba dance. This representation is based on a video analysis of body movements, carried out from the viewpoint of the musical meter. Our method provides the periods, a measure of energy and a visual representation of periodic movement in dance. The method is applied to a limited universe of Samba dances and music, which is used to illustrate the usefulness of the approach.

1. Introduction

In any human societies, dance and music appear as significant components of human expression [1, 2]. Samba represents the most recognizable Brazilian dance, music, social event and way of life. Our analysis focuses on the Samba-no-pé dance, a specific dance style, that is maybe the most recognizable dance style that populates the imaginary of Samba culture among Brazilians and also non-Brazilians [3].

1.1 Samba and multimodality

Samba performers, dancers, musicians, masters or listeners tend to understand Samba as a phenomenon in which music and dance are intrinsically related with each other. However, what is the actual knowledge about the structure of both domains?

Samba music structures involve a polymetric rhythmic texture, with large and strongly syncopated rhythmic formulas in the mid spectra. An ambiguous binary bar is supported by bass lines that accentuate the second beat and often damp first beat notes [4, 5]. Tatum lines in 1/4 beat onsets are concentrated in the high spectra and are subjected to syncopated accents and micro-time deviation profiles [6, 7].

Samba dances have been poorly described, such as in [8] and [9], who also take the musical context into account. Sodré [5] suggests that movement *pushes* the ambiguity of syncopation out of the music. Browning [9] refers to this effect as a musical capacity to provoke *hunger* of movement. All these descriptions acknowledge that body movement has a special place within Samba culture: body movement or dancing seems to be necessary to support the cultural form. It is a condition for music understanding. But how can these hypotheses about music and dance be tested?

The challenge to analyze highly multimodal contexts such as Samba is to develop tools that allow the penetration into the shared structure of music and dance. In this study we propose a method that looks for shared elements in dance and music at the metrical level.

In the following sections, we briefly describe the background of our study. In the second and third section, we summarize the analysis method and describe the representation of dance gestures. In the last sections, we demonstrate how our method leads to musically relevant analysis of dance gestures.

2. Analysis of multi-modal contexts

Samba dance was studied in detail by [10], [11] and [8], who provided us with different types of insightful qualitative analyses on Samba dance, based on subjective methods. Such methods can be complemented with new opportunities that describe gestures with the help of media technologies and computational tools. Computational approaches to dance analysis offer a promising new field for the exploration of dance because they provide access to levels of dance analysis that are difficult to examine by means of the traditional phenomenological methods.

Relevant work in this area includes that of Guedes [12] and Borchers [13] who developed real-time rhythmic analysis of dance respectively using video and sensors. Camurri [14] developed several video-based methods to extract and manipulate dance features in real-time, using the Eyesweb platform. Shiratori and colleagues [15] used musical rhythms to segment and classify dance motions. Jensenius [16] proposed to use a technique called *motiongrams* to visualize structural aspects of dance movements caught on video. Matsumura et al. [17] studied skill acquisition of Samba musicians analyzing data from accelerometers that are attached to the body of dancers. However, the computational approaches that have been developed so far could profit from a multi-modal viewpoint to observed intrinsic relationships between dance gesture and music structures.

In this paper we propose a method that is based on a similar objective (third person) analysis of body movement but which considers Samba dance from the viewpoint of musical meter. Our multi-modal approach allows a proper representation of the basic spatial patterns that underlay the repetitive movement of different body parts along dances. To search for periodicities in the dance movements, we apply the Periodicity Transforms from Sethares & Staley [18], which we use in a proper heuristics that favours metrical structures. By analyzing the evolution of the shape of repetitive dance patterns we were able to grasp meaningful behavior of dancers engagement in meter. This allows us to make straightforward musically relevant descriptions that aim to clarify how dancers may perform their strategies of re-enactment.

3. Methodology for dance analysis

Our analysis method consists of three steps. The first step is to perform movement tracking of the dance in order to obtain the trajectory of the dance movements. In the second step, we perform a decomposition of periodic movements that are related to music periodicities (meter). In the third step, we use the results to analyze the evolution of localized dance movements from the viewpoint of the found metric periodicities.

3.1 Movement tracking and dataset

Two Brazilian professional dancers and teachers performed a total of 6 homogeneous and simple dance excerpts. 3 different Samba music stimuli were chosen from their own repertoire. The trajectories of 9 body points in the visual 2 dimensional plane of video were determined using manual movement tracking [see 19, 20]. It consists of marking the position (horizontal/vertical pixel position) of a desired visual element for each video frame. In this study, 9 points were identified and marked: nose, left shoulder, left rip, hands (left and right), knees (left and right) and feet (left and right). The procedure shown in Fig. 1 was performed using a specific patch in Eyesweb [14] platform.

Fig. 1. Frame-by-frame manual tracking. Pixel positions are marked with the mouse using visual identification .

A set of 18 vectors (2 x 9 body part) was generated with the same temporal definition and spatial resolution of the DV video format (30 fps and 720 by 480 pixels).

3.2 Movement analysis

We use the Periodicity Transforms (PT) to find the most relevant dance patterns, relying on periodicity information from the metrical layers. The metrical layers, or the periods of the music meter grid, were here defined by multiples and divisions of detected beat period in the musical audio excerpts. It is expected that relevant periodic movements will synchronize with these periods. As such we use a multi-modal approach to dance analysis, based on musical meter information.

3.2.1 Metric Layers

Like most dance forms, Samba dance is rhythmically ordered to music, and Samba music is rhythmically grouped and organized by musical meter. Although Samba is rooted in non-western traditions, its evolution combined well-known elements of western music, such as isochronous beats, homophonic texture and a binary meter with elements of African music: rhythmic priority, polymetric lines, syncopation and rhythmic ambiguity [3, 21-24].

The beat markers for each musical excerpt used in the dances were extracted using manual inspection of beat tracking available in the software Beatroot [25]. To compute the bandwidth of the metric layer's periods, we used the mean beat period multiplied by the following metric rule: 0.25, 0.33, 0.5, 0.66, 1, 1.5, 2, 3, 4.

We incorporate knowledge from music into the heuristics to provide a better lens to the dance phenomena. Our method uses meter and BPM information to select relevant information in the movement, mirroring the elements that dancers use to perform dances in synchronization with music structure.

3.2.2 Analysis of periodic patterns

The Periodicity Transforms (PT) basically searches for periodic events in the data. The PT was introduced by Sethares & Staley [18] and further applied in different fields of study such as rhythm analysis [26], analysis of brain waves [27], video and audio integration [28], data mining [29] and bioinformatics [30]. The core element of the algorithms is to decompose the signal in periodic sequences by projecting the given list of periods onto a "periodic subspace". The mathematics of the "projection" procedure is based in a modified form of the *Projection theorem* from Luenberger [31] and detailed in [18, p. 2956]. Sethares' implementations manipulate the projections of each periodicity, subtracting it from the signal, and then repeating it again using the next periodicity in the list over the residue. Implementations of PT provide an output of (i) the period of each repetition, (ii) the measure of energy (norm) extracted from the original signal by each periodicity, and (iii) the periodic basis (waveform) itself [for more information see 18].

Unlike other methods such as Fourier or Wavelet transforms, the PT finds their own bases, and these bases are non-orthogonal. It implies that different orders of projections and subtractions from the signal lead to different results. In other words, there is a preliminary list of periods, and its configuration of order and elements strongly influences the results. Like autocorrelation, PT offers a good definition in low frequencies (large periods), which proves to be more advantageous than Fourier methods. The latter are linear in frequency and lose definition in low frequencies. However, unlike autocorrelation, the PT approach allows the extraction of both the temporal aspect (duration of the beat) and the spatial aspect (the pattern between two beat points).

In our implementation of the PT concept, we use the dependency of PT on the configuration of the initial list of periods to develop a heuristics that mirrors the priorities of the dancer while moving along Samba music. Two simple priorities are applied in our heuristics: (1) that dancers would prefer movements that are repeated periodically around the musical meter and (2) that large movements would be more important that small ones.

Fig. 2 describes this process. The algorithm projects all periodicities until a given N number of samples (normally ½ of total samples of the signal) and filter the periods whose energy peak is above a threshold *th*. In the sequence, it selects only the periods whose peaks that are close to those of the music metric (they may reflect these layers and provide synchronized or counterpoint movements in relation to musical sound). Finally, the algorithm projects the periods of these powerful periodicities in a descendent order of energy, which aims to provide the "best" periods for metrical movements. These are the main points on which the so-called "Best-Route" algorithm relies. The diagram displayed in Fig. 2 subsume the main steps of the algorithm heuristics [for a more complete description of the method, see 32].

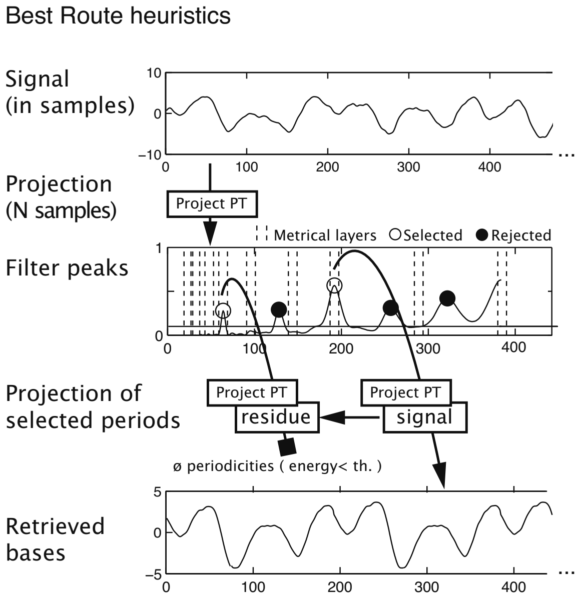


Fig. 2. Diagram of the Best-Route heuristics.

If there are strong periodicities that are not synchronized with music meter such periodicities will not be selected. Ambiguous meter movements will be maintained depending on the domain knowledge imposed in the construction of the metrical grid. The lack of orthogonality allows irrelevant periodicities to interfere in the result. It also permits that the same movement can be seen from different perspectives, depending on the interpretation of the musical meter. Fig. 3 demonstrates the simulation of deviations applied to the metrical grid and the consequent elimination of non-metrical periodicities.

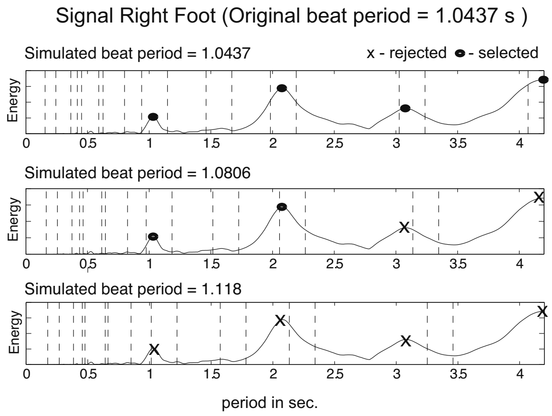


Fig. 3. Simulation of deviations in the metrical grid.

3.3 Spatial representations

By applying the *Best-Route* algorithm, we obtain the periods, energy measures (norm) and waveforms of the most powerful periodic patterns that match with the metrical grid.

However, periodicities from one dimension in time deliver incomplete information to our analyses because they do not re-integrate the two-dimensional vectors from a planar perspective of the video. An immediate solution for this problem would be to conjugate the periodicities found in the complementary dimension (in the last case, the vertical component). However, it is not guaranteed that the analysis of two complementary dimensions in a 2D movement plane shows strong periodicities in the same metrical layer. We solve this problem by "forcing" the PT projection of the missing relevant period onto the complementary dimension. Thus, if there is only one powerful periodicity for a given metric layer, we project its complementary dimension/period onto a periodic subspace without further manipulation of the list. By looking at all periods of the metric grid, and projecting all missing periods, we are able to output hypothetical 2D projections of periodic movements related to each metric layer. Relevant periodicities (powerfully evident and metrically relevant) are indicated by means of the Best-Route algorithm. Fig. 4 describes this process.

3.3.1 Decomposition of movement

The outcome of the Best-Route heuristics can provide 3 qualities of spatial representations:

- Bidimensionally evident (B), if the best periodicities are found in both vertical and horizontal dimensions at the same metrical level.
- Unidimensionally evident (U), if only one dimension (Uh-horizontal or Uv-vertical) was chosen as relevant by the algorithm.
- Evident periodicities (N), if both components are missing and the projection must be projected using the metrical level period in both vectors.

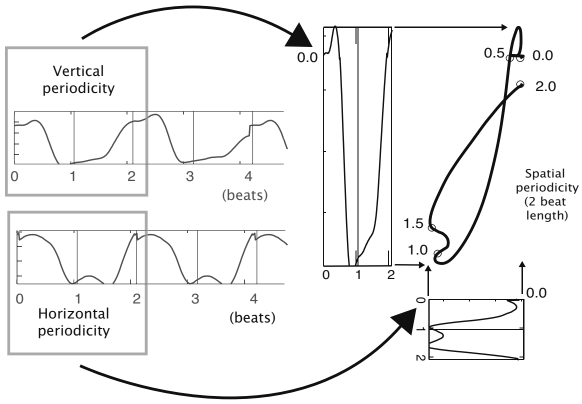


Fig. 4. Reintegration of movement in spatial descriptions.

As displayed in Fig. 5, the most relevant metrical levels show large visual magnitudes and are directly related with powerful periodicities found in one (B) or both dimensions (U) by the Best-route algorithm. This visualization also shows how the hypothesized movement may be evidenced by each metrical layer, although not all metric layers may be perceptually evident. Note that Bidimensionally evident (B) patterns are not necessarily the most evident ones, because secondary periodicities (weak) can be found at the same metrical layer.

This method provides not only a general basis of repetition that allows us to rebuild the trajectories in a periodic representation, but also a systematic connection with periods and relative energy measures of these movements. It surpasses frequency definitions of Fourier methods in low frequencies and shows insightful representations resulted from the projected periodicities (basis). The shape in the metrical level 2.00 seen in Fig. 5, for example, shows how the left hand movements are repeated on the video screen at every two beats during the dance. Each of these patterns can also be segmented into time steps, which provides us with detailed information about the evolution of the movement in each phase of the metric engagement. Fig. 5 shows one example segmented into half-beat steps.

By using this approach we can describe the dance form in terms of repetitive metric patterns. Spatial descriptions can be used to observe how repetition in dance movements shows the deployment of bodily engagement in Samba dance choreography, along with Samba music.

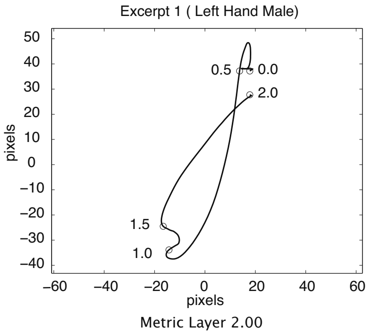


Fig. 5: Pattern of the left hand (male dancer, excerpt 1), in the metrical level of 2 beats (Bar). Marked points indicate 0.5 beat steps.

4. Results

In this study we show that repetition gestures in dance examples can be represented in a proper and meaningful way, using the above mentioned heuristics. For an easy recognition of indications, right and left sides will be referenced here as the viewpoint of the observer (inverted in relation to the dancer).

Fig. 6 shows descriptions grouped across metrical layers, plotted against one frame of its original video. This visualization provides insightful maps of gestures that visually cluster characteristics of movement patterns. Although the horizontal position is moved to separate each metrical layer (signalized in the bottom part) the reader can rely on the vertical position shifting the patterns horizontally to grasp the behavior of the body parts. Fig. 7 shows a gesture grid that compares the movement of the Right Hand in all metrical layers (columns) from all dancer excerpts (rows). Grouped gestures offer interesting visualization cues that make similarities and dissimilarities evident. Note that patterns from excerpts 1-3 (displayed in rows) show those of the male dancer, while excerpts 4-6 show those of the female dancer. Fig. 8 a-b shows 2 detailed descriptions, which demonstrate how gestures can be analyzed by its evolution in metric engagement (time).

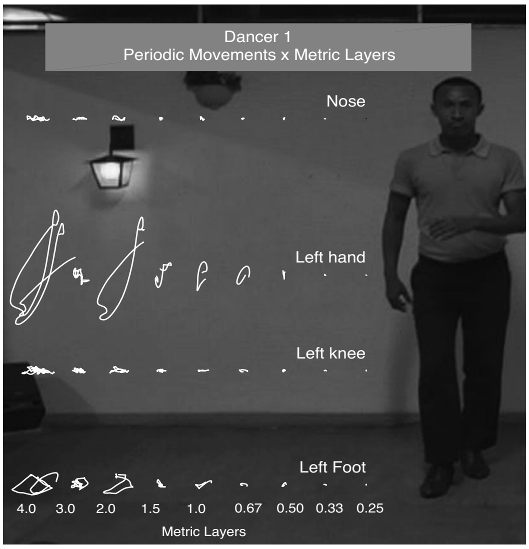


Fig. 6. Video frame of dancer, superimposed by groups of patterns by metrical layer.

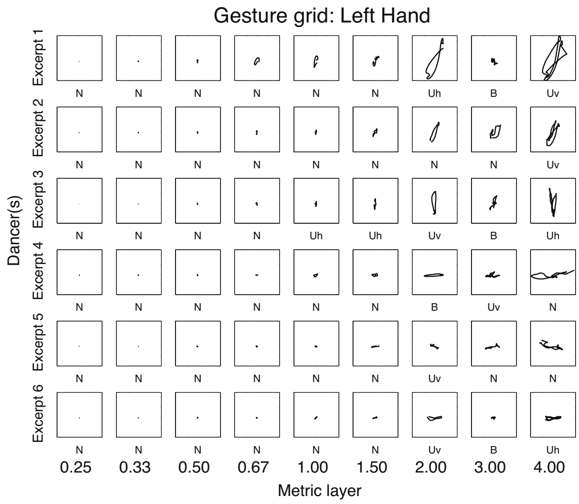


Fig. 7. Grouped patterns representing left hand redundant patterns.

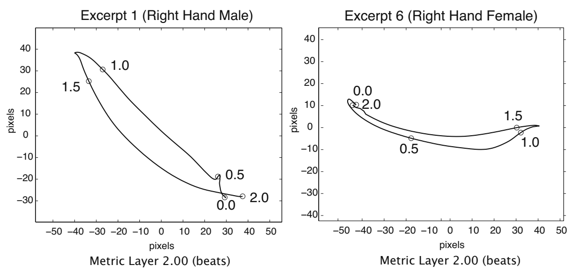


Fig. 8 a-b. Two examples of male and female left hand movement in the 2 beat metrical layer. Points marked indicate each annotated segment in 0.5 beat steps.

5. Discussion

It is questionable whether it is ecologically consistent to “encapsulate” the movement in arbitrary 2 axes (vertical and horizontal) while the dancer does not build artificial axes to perform her/his movement. However, the temporal marks such as beat and periodic lines are good musical indications that show links between movement and meter and musical time.

The grouped descriptions shown in Fig. 6, 7 and 8 provide interesting visualizations of how metric layers 2 and 4 (bar and double bar) are stressed in Samba dance engagement. Such characteristics were verified in all dances (1-6) by looking at the amplitude of the 7th and 9th patterns (2 and 4 beat metrical layers). Individual differences in these 6 dances interestingly emerge by comparing the patterns in rows. The grouped visualization in Fig. 7 shows that the male dancer (rows 1-3) tends to move his hand in vertical patterns, while the female dancer (rows 4-6) follows horizontally oriented gestures.

A more detailed analysis displayed in the Fig. 5 shows that thetic movement of the male dancer is concentrated on the extremity of these gestures. The evolution of his movement oscillates in fast changes before the beat marks (1.0 and 2.0). This suggests a strong gesture intention in the direction of beat indications, also verified in Fig. 8 a and b. Subtle movements mark 1.5 points, which could suggest a polymetric engagement, also denoted by emergence of patterns in the 3.00 beat layers in Fig. 7. Hands seem to stress beat points using displacement stops. Fig. 8 (a and b) shows an interesting comparison. Although the direction of gestures differ and the hands are in contra-phase (see 0.0 labels along patterns), both dancers change the direction between 1.0 and 1.5 beat points, which signalizes coherence of gesture forms between dancers. The female left-hand gestures seem to move more horizontally, in a pendulum-like form, while the male dancer tends to perform diagonal movements.

At the top of our observations, the comparison between video excerpts and representations suggests that our method provides relevant information about the dances. Successful sonifications of these patterns in Samba music sequences found by Naveda and Leman [32] also reinforces the extent of the method.

6. Conclusion

The representation, characterization and segmentation of repetition in movement by its relevant musically structures (meter) can provide a basis for a useful methodology in dance analysis, especially for dance forms. Our results show that the coupling between body and sound, culturally verified in the ecological domain, can be extended to the analytic approaches, generating relevant results. The extent of the method ranges from the detection of periods of repetition to the deployment of subtle characteristics of the movement into fractions of the musical beat. The method seems to confirm the link between Samba dance forms and musical meter, as proposed by several earlier musicological and anthropological studies.

7. Acknowledgements

This research is supported by a grant from University of Ghent, and partial support by CAPES/Brazil. Many thanks to the dancers Lorena Anastácio and Vagner Santos who provided insights and significant dance performances for this study.

References

[1] J. L. Hanna, R. D. Abrahams, N. R. Crumrine, R. Dirks, R. V. Gizycki, P. Heyer, A. Shapiro, Y. Ikegami, A. L. Kaeppler, and J. W. Kealiinohomoku, "Movements Toward Understanding Humans Through the Anthropological Study of Dance [and Comments and Reply]," *Current Anthropology,* vol. 20, pp. 313-339, 1979.

[2] A. Grau, "Sing a Dance. Dance a Song: The Relationship between Two Types of Formalised Movements and Music among the Tiwi of Melville and Bathurst Islands, North Australia," *Dance Research: The Journal of the Society for Dance Research,* vol. 1, pp. 32-44, 1983.

[3] M. E. Mariani, "African influences in Brazilian Dance," in *African Dance*, K. W. Asante, Ed. Asmara: Africa World Press, 1998, pp. 79-98.

[4] C. Sandroni, *Feitiço decente: transformações do samba no Rio de Janeiro, 1917-1933*: Jorge Zahar Editor: Editora UFRJ, 2001.

[5] M. Sodré, "Samba, O Dono do Corpo," *Rio de Janeiro: Codecri,* 1979.

[6] M. Wright and E. Berdahl, "Toward Machine Learning of Expressive Microtiming in Brazilian Drumming," in *International Computer Music Conference*, New Orleans, 2006.

[7] F. Gouyon, "Microtiming in “Samba de Roda”—Preliminary experiments with polyphonic audio," in *XII Simpósio da Sociedade Brasileira de Computação Musical* São Paulo, 2007.

[8] M. E. Mariani, "A Portrayal of the Brazilian Samba Dance with the Use of Lab Analysis as a Tool for Movement Analysis," University of Wisconsin, Madison, 1986.

[9] B. Browning, *Samba: Resistance in Motion*: Indiana University Press, 1995.

[10] M. A. C. Giffoni, *Danças folclóricas brasileiras e suas aplicações educativas*. São Paulo: Edições Melhoramentos, 1964.

[11] J. Chasteen, "The prehistory of Samba: Carnival Dancing in Rio de Janeiro, 1840-1917," *Journal of Latin American Studies,* vol. 28, pp. 29-47, Feb. 1996.

[12] C. Guedes, "Extracting Musically-Relevant Rhythmic Information from Dance Movement by Applying Pitch-Tracking Techniques to a Video Signal," in *Proceedings of the Sound and Music ComputingConference SMC06*, Marseille, France, 2006, pp. 25-33.

[13] U. Enke and J. Borchers, "Rhythmic Analysis of Dance Movements using Acceleration-Onset Times," in *Dept. of Computer Science* Aachen: RWTH Aachen University, 2006.

[14] A. Camurri, S. Hashimoto, M. Ricchetti, A. Ricci, K. Suzuki, R. Trocca, and G. Volpe, "EyesWeb: Toward Gesture and Affect Recognition in Interactive Dance and Music Systems," *Computer Music Journal,* vol. 24, pp. 57-69, 2000.

[15] T. Shiratori, A. Nakazawa, and K. Ikeuchi, "Detecting dance motion structure through music analysis," *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on,* pp. 857-862, 2004.

[16] A. R. Jensenius, "Using motiongrams in the study of musical gestures," *In Proceedings of the 2006 International Computer Music Conference,* pp. 6-11, 2006.

[17] K. Matsumura, T. Yamamoto, and T. Fujinami, "Analysis of Samba Dance Using Wireless Accelerometer," *Nihon Kikai Gakkai Supotsu Kogaku Shinpojiumu, Shinpojiumu Hyuman, Dainamikusu Koen Ronbunshu,* vol. 2006, pp. 216-221, 2006.

[18] W. A. Sethares and T. W. Staley, "Periodicity transforms," *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on],* vol. 47, pp. 2953-2964, 1999.

[19] N. Rossini, "The Analysis of Gesture: Establishing a Set of Parameters," *Gesture-Based Communication in Human-Computer Interaction: 5th International Gesture Workshop, GW 2003: Genova, Italy, April 15-17, 2003: Selected Revised Papers,* 2004.

[20] M. Clayton, R. Sager, and U. Will, "In time with the music: The concept of entrainment and its significance for ethnomusicology," *ESEM Counterpoint,* vol. 1, pp. 1–82, 2004.

[21] C. Sandroni, "Dois sambas de 1930 ea constituição do gênero: Na Pavuna e Vou te abandonar," *Cadernos do Colóquio,* vol. 1, 2001.

[22] G. Kubik, "Analogies and Differences in African-American Musical Cultures across the Hemisphere: Interpretive Models and Research Strategies," *Black Music Research Journal,* vol. 18, pp. 203-227, 1998.

[23] P. Fryer, *Rhythms of Resistance: African Musical Heritage in Brazil*: Pluto, 2000.

[24] J. J. d. Carvalho, *Afro-Brazilian Music and Ritual [s]*: Duke-University of North Carolina Program in Latin American Studies, 2000.

[25] S. Dixon, "BeatRoot 0.5.6," 2007.

[26] W. A. Sethares and T. W. Staley, "Meter and Periodicity in Musical Performance," *Journal of New Music Research,* vol. 30, pp. 149-158, 2001.

[27] P. C. Lo and J. S. Leu, "Quantification of Pseudo-Periodicity of Alpha Rhythm in Meditation EEG," *Journal of Medical and Biological Engineering,* vol. 25, pp. 7-13, 2005.

[28] S. Ravulapalli and S. Sarkar, "Association of Sound to Motion in Video using Perceptual Organization," *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on,* vol. 1, 2006.

[29] Y. Cai, R. Stumpf, T. Wynne, M. Tomlinson, D. S. H. Chung, X. Boutonnier, M. Ihmig, R. Franco, and N. Bauernfeind, "Visual transformation for interactive spatiotemporal data mining," *Knowledge and Information Systems,* vol. 13, pp. 119-142, 2007.

[30] A. K. Brodzik, "Quaternionic periodicity transform: an algebraic solution to the tandem repeat detection problem," *Bioinformatics,* vol. 23, p. 694, 2007.

[31] D. G. Luenberger, *Optimization by Vector Space Methods*: Wiley-Interscience, 1969.

[32] L. Naveda and M. Leman, "Sonification of Samba dance using periodic pattern analysis," in *Artech 2008* Porto: Portuguese Catholic University 2008.

SYSSOMO: A Pedagogical Tool for Analyzing Movement Variants Between Different Pianists

Aristotelis Hadjakos Erwin Aitenbichler Max Mühlhäuser
Telecooperation Group, Department of Computer Science, TU Darmstadt, Germany
E-mail: {telis, erwin, max} @tk.informatik.tu-darmstadt.de

Abstract

The visualization of arm movement can provide valuable additional information for piano teaching. However, movement visualizations are often difficult to understand for musicians. SYSSOMO uses score-following to synchronize two performances of the same piece in order to overlay motion data, MIDI, and video. Using this technique, the user can easily compare the differences between the performances of a teacher and a student.

To capture arm movement, sensors have to be unobtrusive for good user acceptance, and lightweight to minimize independent movement of the sensor. We have developed and built sensing hardware, called MotionNet, to capture arm movements. MotionNet is composed of a set of sensor units and a host unit, which communicate via CAN. This architecture helps us to reduce size and weight of the sensor units.

1. Introduction

The movement of the arm plays an important role in piano technique. The arm has to bring the finger, which executes the next touch, in an optimal position to strike the key. Arm movement can also be used to execute a touch while the finger remains slightly fixated to transduce the movement to the key.

We were able to identify piano playing variants that are distinguishable on gyroscope and accelerometer sensors attached at different positions of the player's arm. E.g., it is possible to distinguish variants of tremolo playing. We have published these result in [7].

In our experience, graphs of movement data are often difficult to understand for musicians. It is our approach to overlay two performances, for example the performance of a teacher and the performance of a student, so that users can easily see the differences.

The SYSSOMO system (SYnchronization through Score-following of SOmatic MOtion data) visualizes data captured by gyroscopes and accelerometers, which are attached to the user's arm. SYSSOMO has the ability to synchronize two performances and to overlay the motion data visualization. To support different learning scenarios, including distance learning, SYSSOMO provides further features: additional modalities, i.e., audio, MIDI, piano roll, and video, efficient storage, and annotation of performances.

The rest of this paper is organized as follows. In section 2 we discuss related work. The design of MotionNet is presented in section 3. In section 4, the synchronization mechanism of SYSSOMO and the additional features are discussed. A usage example is presented in section 5. Finally, the paper is concluded in section 6.

2. Related Work

2.1 Piano Teaching Systems

A variety of piano teaching systems have been proposed, including the famous Piano Tutor [4], the fingering suggestion and animation system by Lin et al. [9], the scale evaluation system by Akinaga et al. [1], the exercise generation and feedback system by Kitamura and Miura [8], the analysis and exercise generation system by Mukai et al. [12], the piano duo support system by Oshima et al. [14], and our recent sonification system [6]. SYSSOMO is a piano teaching system that uses performance visualization to provide an additional information source for piano tuition.

2.2 Performance Visualization Systems

Performance visualization systems process input data that is captured during the performance and visualize certain aspects of the performance.

Performance visualization systems need not to model "correct play" and are therefore not limited to the musical learning scenarios that the developer has foreseen. Because of the high level of detail, which is often provided by performance visualization systems, the interpretation of the visualizations can be difficult and require an advanced user. However, performance visualization systems can provide information that is otherwise unaccessible and can therefore be a valuable additional information source for instrument tuition.

2.2.1 Motion Data-based Performance Visualization Systems

Montes et al. [10] visualized surface EMG signals from relevant muscles to teach thumb touches. The EMG signals show different muscular activation patterns when comparing professional pianists with non trained persons.

Riley uses a system to capture MIDI, video, and surface EMG of finger muscles for teaching the piano [15, 16]. The system visualizes MIDI data as piano roll to visualize musical issues, video for frame-to-frame motion analysis, and surface EMG signals to identify unnecessary tension.

Mora et al. [11] developed a system that overlays a 3D mesh of a suggested posture on a video of the student's performance. The student can visualize the differences and adopt the suggested posture. To generate the 3D mesh, the posture of a professional pianist was recorded using motion capturing.

The 3D Augmented Mirror (AMIR) system [13] is a performance visualization system for string instruments. Markers that are attached to the instrument, the bow, and the player‘s body are tracked with a VICON. The AMIR provides several visualizations including a 3D rendering of the performance.

The visualization systems of Bouenard et al. [3] is a performance visualization system for timpanists. Markers that are attached to the drumstick and the player‘s body are tracked with a VICON. The system shows an animation of a 3D virtual character and provides several visualizations of kinematic and dynamic cues. Kinematic cues include plots of position, velocity, and acceleration as well as combinations of position-velocity and velocity-acceleration.

In contrast to other motion data based performance visualization systems, SYSSOMO provides the ability to synchronize two performances so that they can be compared easily.

2.2.2 MIDI-based Piano Performance Visualization Systems

The pianoFORTE [18] system uses MIDI input and visualizes tempo, articulation, and dynamics of the performance. The visualization is based on a music score, which is annotated to reflect dynamics and articulation of the performance. Tempo is visualized by a speed-o-meter. Missed and wrong notes are marked in the score.

The MIDIATOR [17] uses MIDI data to visualize note timings and note volumes. Note timings are shown in a piano roll representation and allow the analysis of rhythm, articulation, and tempo. It is possible to superimpose two performances. However, the two performances have to be played in exactly the same tempo.

The practice tool for pianists by Goebl and Widmer [5] generates visual feedback from MIDI input in real-time. The practice tool finds reoccurring patterns by autocorrelation and visualizes them. By placing successive patterns, represented over each other in a piano roll fashion, timing deviations can be seen. Additionally, the practice tool provides several other visualizations including a visualization of automatically extracted beats, a chord timing visualization and a piano roll overview.

SYSSOMO uses score-following to automatically synchronize two performances and can therefore, differently than the mentioned MIDI-based performance visualization systems, superimpose performances with tempo variations.

3. Sensors

We designed and built a custom sensing platform for SYSSOMO. Our sensors capture arm movements by inertial measurement, based on gyroscopes and accelerometers. MotionNet was developed according to the following requirements. We wanted our sensor system to be:

Unobtrusive: As the sensors have to be worn during piano performance, the sensors have to be as unobtrusive as possible to be accepted by users.

Lightweight: Since sensors can never be fixed perfectly to the arm, a low weight is essential. Otherwise the mass inertia of sensors would lead to uncontrolled, independent movements of the sensor itself, which has a negative impact on the accuracy of measurements.

Real-time capable: To support sonification of movements and live remote performance and teaching, the communication channel between the sensors and the host must have a high bandwidth and a low latency.

Configurable: The user should be able to balance wearing comfort and level of detail of the captured motion data. It should be possible to add or remove sensors easily.

Low cost: A large part of the intended target users, e.g., small musical schools or pianists with no professional ambition (yet) should be able to afford the system as well.

Because we could not find any system on the market that meets our requirements, we started building our own hardware. To date, wireless sensor platforms (e.g., Crossbow [19]) are in general not suitable for our application. Wireless communication bandwidth is too little, media access control (MAC) only scales up to a few nodes, and the message delay is significantly higher compared to wired sensor networks. In addition, such sensors must be powered by a battery that adds weight. Off-the-shelf inertial sensors (e.g., InertiaCube [20], Xsens [21]) are very expensive, i.e., about ten times more expensive than a sensor of our system. The InertiaCube is not based on a bus, which raises the question of how to connect many sensors to a host. The closest system is the Xsens MTx Xbus system. It would have been usable for our application, but was out of our budget. When sampling at 100 Hz, the Xsens sensors compare to our system as follows:

Feature / System	Xsens MTx	MotionNet
Max. number of sensors	10	80
Weight	30g	10g
Magnetic field sensor	yes	no
Power consumption	360mW	170mW
Cost (approx.)	EUR 1.750,--	EUR 150,--

3.1 Hardware

The hardware of the MotionNet system consists of an arbitrary number of *sensor units* and a *host unit*. All units are interconnected via a CAN bus (ISO 11898). Because CAN uses up to 29 bits for addressing, the maximum number of sensors is only determined by the total bandwidth of the bus (1 Mbit/s) and physical deployment considerations.

3.1.1 Host Unit

The *host unit* is used to connect a computer to the sensors. It powers the bus and all sensors and bridges from the CAN bus to RS232 (up to 921.6 kbps).

When the system is powered up, the host broadcasts a discovery message. Each sensor unit receiving this message sends a reply back to the host containing its identifier. This allows the host to create a list of sensors present on the bus.

During measurement, the host unit broadcasts sensor trigger messages at a rate of 100Hz. This allows to synchronize all sensors. The host unit then waits until all discovered sensor units have replied with their sensor values and sends the collected result to the computer via the serial interface.

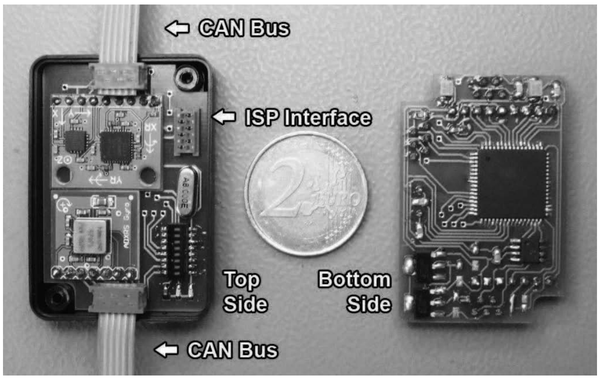


Figure 2: Sensor unit

3.1.2 Sensor Unit

Each sensor unit (Figure 1) has a unique identifier, which unambiguously describes the position where it has to be worn by the user. This identifier is configured with hardware DIP switches.

The central component of the sensor unit is an Atmel AT90CAN128 microcontroller. 3D acceleration is measured with an ADXL330 sensor and 3D rotation is measured with a combination of two gyroscope sensor chips: an IDG300 for the *x*- and *y*-axes and an ADXRS300 for the *z*-axis.

The sensor unit therefore provides six measurements (6-DOF) in total (Figure 2).

3.2 Mounting

Possible positions for the sensor units are the back of the hand, the wrist, and the upper arm. The user can choose which combination of sensor positions suits her best. We used a combination of Velcro fasteners and rubber bands to fixate the sensors to the arm.

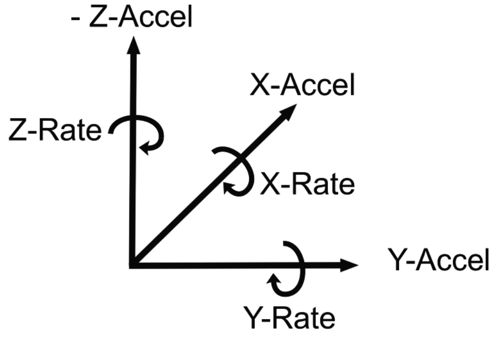


Figure 1: Measured axes

4. SYSSOMO

The following scenarios should be supported by SYSSOMO:

Single user: The user wants to analyze her playing movements with SYSSOMO. She might want to compare an old recorded performance with a new performance. Alternatively, she might want to compare her performance with the performance of her teacher.

Two users: Two users want to compare their performances and study the differences of their playing movements.

Distance learning: The student records motion data and sends it to her teacher over the Internet. The teacher compares the recording of the student to a reference recording, which the teacher has generated beforehand.

The following requirements can be derived from the described usage scenarios:

Visualization: Motion data should be displayed.

Comparison: The system should support the users to compare different performances.

Persistence: It should be possible to make recordings of the motion data of a performance and to load recordings at a later time.

Multimedia: The system should also support audio, video, and MIDI data.

Encoding: To allow transmission over the Internet, the size of the data to be transmitted has to be kept small.

Annotation: The system should allow to annotate motion data.

4.1 Visualization

SYSSOMO has four sources of input:

- Gyroscope and accelerometer data from the MotionNet
- MIDI data (we used a Kawai K-15 ATX MIDI-enabled upright piano)
- Video data (we used a Genius Slim 1322AF webcam)
- Audio data (we used an external microphone)

SYSSOMO provides the following visualizations (Figure 3):

- Motion visualization of the gyroscope and accelerometer data
- Piano roll of the MIDI data
- Video

The motion visualization shows the graphs of the gyroscope and accelerometer data of a sensor node. The zero-line is marked along with +/-50°/sec markings for the gyroscope and +/-1g for the accelerometer signal. The user can display several motion visualizations simultaneously for the different sensor units attached to the arm. For the interpretation of the graphs, it can be important to know when a note was played. Therefore, we included dots that represent the played notes at the bottom of the motion visualization. These dots also help the user to switch between the motion visualization and the piano roll.

4.2 Comparison and Persistence

SYSSOMO allows the user to save recordings of the performance. This includes the motion data, MIDI, audio, and video data. When a saved performance is opened, the user can control the visualization playback of the motion data, MIDI, audio, and video data with a control similar to a tape deck control (play, pause, fast forward, fast backward). The user can also hear the audio or the MIDI recording.

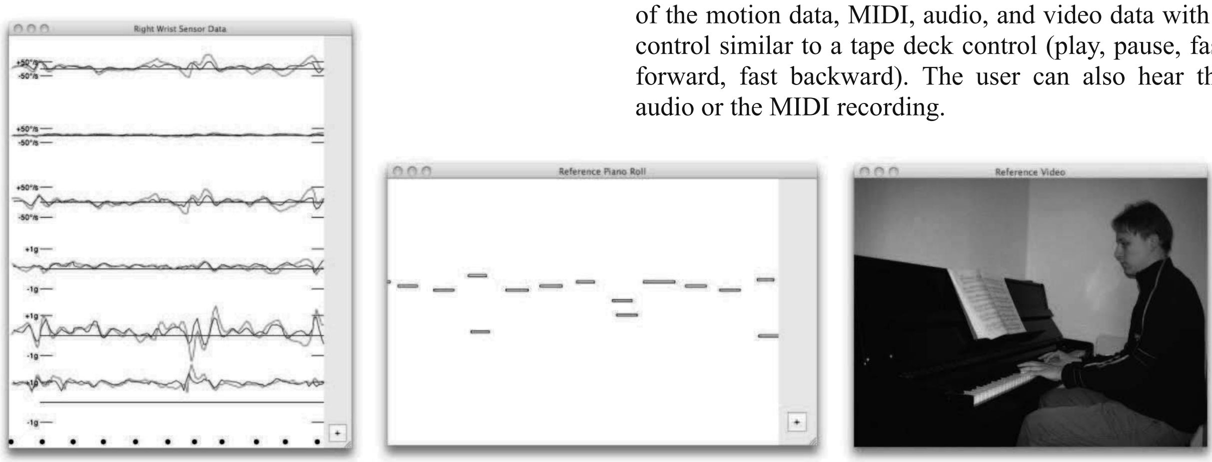


Figure 3: Motion data visualization, piano roll, and video

SYSSOMO provides the ability to align two performances of the same piece, facilitating their comparison. SYSSOMO distinguishes between the **reference performance** and the **dependent performance**: While the reference performance stays unchanged, the dependent performance is modified so that reference and dependent performance are synchronized.

To modify the dependent performance, SYSSOMO needs to compute a time map, which translates from dependent performance time to reference performance time. To this purpose, SYSSOMO aligns the MIDI data of the performances. SYSSOMO synchronizes the performances using Bloch's and Dannenberg's polyphonic score follower [2]. The algorithm matches MIDI events of the dependent performance to MIDI events of the reference performance. The timestamps of the matched MIDI events provide synchronization points between dependent and reference performance. The time map is completed by using linear interpolation between the synchronization points. Future developments could include audio-based score-following to support acoustic pianos without MIDI interface. By using the time map, SYSSOMO can synchronize motion data, MIDI, and video of the two performances. Future developments could include audio time-stretching of the dependent performance.

When a user plays the same passage at very different speeds, the playing movements usually have different forms, e.g., the size of the playing movements usually reduces with increased tempo. Furthermore, the time scaling of movement data introduces artifacts that are noticeable at very different tempi. The users should therefore normally use SYSSOMO to synchronize performances with comparable tempos.

4.3 Multimedia and Encoding

To enable transmission over the Internet, the size of the stored recording has to be kept relatively small. The data rate of the captured motion data is small compared to the amount of data for audio and video capturing so it did not seem necessary to compress the motion data. For video, the user can choose between H.264 and MPEG4 codecs with different resolutions. For audio, the user can choose between the high-fidelity ALAC codec and the AAC codec, which is available for different bitrates.

4.4 Annotation

The user can annotate the motion data visualization and the piano roll. An annotation can relate to a single point in time or a time interval. An annotation (Figure 4) consists of a label, like “Advice“ or “Question“, a text, and one or two time markers. A time interval is represented by two time markers. If the annotation relates to one point in time, a single time marker is used.

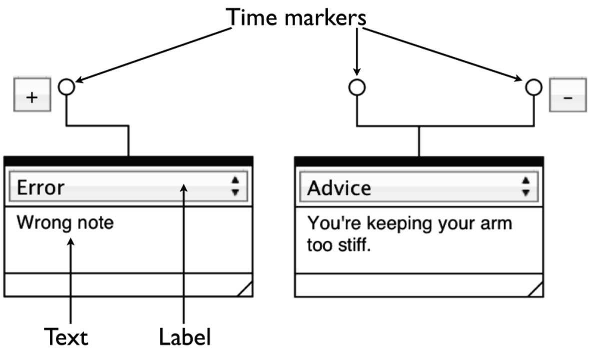


Figure 5: Annotations

5. Usage Example

In the following, we want to demonstrate the interaction with SYSSOMO in a distance learning scenario.

The student wears the motion sensors and captures a performance with SYSSOMO. Using SYSSOMO, she plays back the performance: She hears the music, sees the motion data visualization, the piano roll, and the video. She uses the annotation tools to mark questions and any mistakes that she is aware of. She then sends the annotated performance to her teacher.

The teacher loads the student‘s performance as the reference performance. For comparison, the teacher loads her own performance of the same piece as dependent performance. Hearing the music, the teacher identifies technical problems of the student and uses the motion data visualization to study differences in the movement (see figure 5 for an example). The teacher then uses annotation to give advice. Finally, when the teacher is done annotating the performance, she sends back the annotated performance to the student.

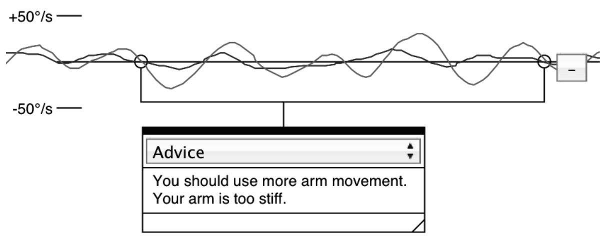


Figure 4: The signal shows the rotation of the arm. The teacher (red) uses more arm movement than the student (blue).

6. Conclusion

The visualization of arm movement can provide valuable additional information for piano teaching. However, movement visualizations are often difficult to understand for musicians. Therefore, SYSSOMO overlays two performances of the same piece so that the user can more easily identify differences of the arm movement. As we have shown previously [7], playing variants are visible on gyroscope and accelerometer signals of sensors that are attached to the pianist‘s arm.

To support different pedagogical scenarios, including distance learning over the Internet, further features were included in SYSSOMO: additional modalities (audio, MIDI, piano roll, video), efficient storage, and annotation of performances.

MotionNet provides SYSSOMO with rate and acceleration measurements of the arm movement. Separating functionality between sensor units and a host unit helps us to reduce size and weight of the sensor units, which is important for user acceptance and to minimize independent movement of the sensor units.

7. References

[1] S. Akinaga, M. Miura, N. Emura, and M. Yanagida. Toward realizing automatic evaluation of playing scales on the piano. In International Conference on Music Perception and Cognition (ICMPC), 2006.

[2] J. J. Bloch and R. B. Dannenberg. Real time computer accompaniment of keyboard performances. In ICMC 1985, 1985.

[3] A. Bouenard, S. Gibet, and M. M. Wanderley. Enhancing the visualization of percussion gestures by virtual character animation. In NIME 2008 Proceedings, 2008.

[4] R. B. Dannenberg, M. Sanchez, A. Joseph, R. Joseph, R. Saul, and P. Capell. Results from the piano tutor project. In Proceedings of the Fourth Biennial Arts and Technology Symposium, pages 143–150, 1993.

[5] W. Goebl and G. Widmer. Unobstrusive practice tools for pianists. In Proceedings of the 9th International Conference on Music Perception and Cognition (ICMPC9), 2006.

[6] A. Hadjakos, E. Aitenbichler, and M. Mühlhäuser. The elbow piano: Sonification of piano playing movements. In NIME 2008 Proceedings, 2008.

[7] A. Hadjakos, E. Aitenbichler, and M. Mühlhäuser. Potential use of inertial measurement sensors for piano teaching systems: Motion analysis of piano playing patterns. In Proceedings of the 4th i-Maestro Workshop on Technology-Enhanced Music Education, 2008.

[8] T. Kitamura and M. Miura. Constructing a support system for self-learning playing the piano at the beginning stage. In International Conference on Music Perception and Cognition, 2006.

[9] C.-C. Lin and D. S.-M. Liu. An intelligent virtual piano tutor. In Proceedings of the 2006 ACM international conference on Virtual reality continuum and its applications, pages 353–356, 2006.

[10] R. Montes, M. Bedmar, and M. S. Martin. EMG biofeedback of the abductor pollicis brevis in piano performance. Biofeedback and Self-Regulation, 18(2), 1993.

[11] J. Mora, W.-S. Lee, G. Comeau, S. Shirmohammadi, and A. E. Saddik. Assisted piano pedagogy through 3d visualization of piano playing. In HAVE 2006 – IEEE Intl. Workshop on Haptic Audio Visual Environments and their Application, pages 157–160, 2006.

[12] M. Mukai, N. Emura, M. Miura, and M. Yanagida. Generation of suitable phrases for basic training to overcome weak points in playing the piano. In Proc. of International Congress on Acoustics, 2007.

[13] K. Ng, T. Weyde, O. Larkin, K. Neubarth, T. Koerselman, and B. Ong. 3d augmented mirror: A multimodal interface for string instrument learning and teaching with gesture support. In ICMI ’07: Proc. of the 9th Intl. Conference on Multimodal interfaces, 2007.

[14] C. Oshima, K. Nishimoto, and N. Hagita. A piano duo support system for parents to lead children to practice musical performances. ACM Trans. Multimedia Comput. Commun. Appl., 3(2), 2007.

[15] K. Riley. New views of piano playing through motion analysis and midi technology. Journal of Technology in Music Learning, 2007.

[16] K. Riley, E. E. Coons, and D. Marcarian. The use of multimodal feedback in retraining complex technical skills of piano performance. Medical Problems of Performing Artists, 20(2), 2005.

[17] S. Shirmohammadi, A. Khanafar, and G. Comeau. Midiator: A tool for analyzing students’ piano performance. Revue de recherche en éducation musicale, 24:35-48, 2006.

[18] S. W. Smoliar, J. A. Waterworth, and P. R. Kellock. pianoforte: A system for piano education beyond notation literacy. In MULTIMEDIA ’95: Proceedings of the Third ACM International Conference on Multimedia, 1995.

[19] Crossbow Technology: Wireless Sensor Networks, http://www.xbow.com/, last visited: 04.07.08

[20] INITION: InertiaCube, http://www.inition.com/, last visited: 04.07.08

[21] Xsens Technologies: Xbus Kit, http://www.xsens.com/, last visited: 04.07.08

The ENABLED Editor and Viewer – simple tools for more accessible on line 3D models

Charlotte Magnusson* Teresa Gutierrez *† Kirsten Rassmus-Gröhn*
(*)Lund University, Lund, Sweden
(*†)LABEIN, Bilbao, Spain
E-mail: charlotte.magnusson@certec.lth.se, tere@labein.es, kirre@certec.lth.se

Abstract

This paper reports on the ENABLED 3DEditor and 3DViewer. The software design is described, and results from user tests with end users are reported. Both the Editor and Viewer are seen to work quite well. It is possible for a developer to quickly start working with the editor. The Viewer was well received by the users who are able to use it to understand an environment, get an overview and locate a specific place on the 3D map.

1. Introduction

Research on presenting Web content through non-visual form is very active especially after the GUI interface has become one of the most prevalent interface types which exist today. Haptic, auditory and multimodal modalities have been used to provide novel methods for people with visually impairments to interact with computer systems. For instance, Ramstein et al [1] have proposed the PC-Access system that offers auditory information, reinforced by the sense of touch to enable blind users to be aware of what happens on the screen and where the pointer is. Rosenberg and Scott [2] have also carried out a study to prove that force feedback can enhance a user's ability to perform basic functions within graphical user interfaces. Using a haptic display in combination with audio feedback is one way to enable access. General guidelines to create and develop haptic applications and models are collected in [3]. Applications making practical use of non-spoken audio and force-feedback haptics for visually impaired people are e.g. applications supporting mathematical display [4], [5] & [6], games [7-9] and audio-haptic maps [7;10;11]. In [12], a CAD application is presented that enables users to create drawings with the help of audio and keyboard. In [13] an audio-haptic drawing program is described. The work described in this paper builds primarily on the previous work reported already in [10] and later in where it was shown that the use of haptic 3D models together with spoken audio tags could be used to allow visually impaired persons to access also quite complex maps. We have also made use of the studies reported in [14] for the design of pan and zoom tools. Thus the design principles are not really new (although the zooming function has been enhanced with audio feedback) – but what is new is the way the developed software can make such environments easier to generate. The work has been part of a European project, ENABLED. One of the objectives of the ENABLED system is to provide an accessible end interface for people with vision difficulties to gain access to various types of graphics on the Web [15].

2. The ENABLED 3DEditor

The ENABLED 3DEditor is a program that can read an X3D 3D model and add haptic surfaces and speech tags to make the model more accessible to persons with visual impairments. The program is implemented making use of the open source api H3D for the haptics and Microsoft Speech - SAPI 5.1 for the speech. It is currently working with the PHANToM Omni, but could be extended for use with the Novint Falcon.

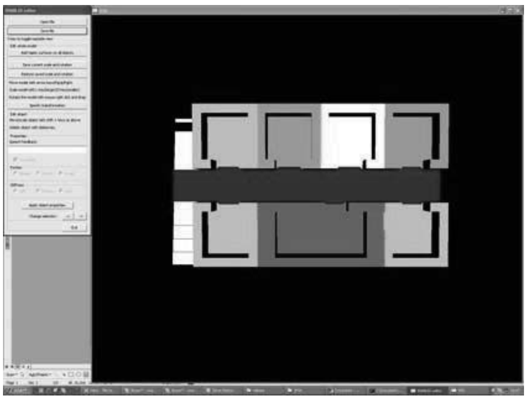


Figure 1. A Screen shot of the ENABLED 3dEditor.

To make the program as easy to use as possible it is quite limited in functionality. One can:

- scale and move the 3D model so that it fits well within the workspace (size and position is a common problem when exporting from e.g 3DStudioMax to put a model on the web)
- add haptic surfaces to the model
- edit friction (slippery, smooth, rough) and stiffness (soft, medium, hard)
- add speech tags to all objects
- switch between viewing objects from the side and from above (figure 1 shows the view of a map environment from above)
- save the results into a file that can be put on the web and which can be viewed using the ENABLED 3DViewer

The 3DEditor is primarily designed for sighted developers (although one blind reference user has also made use of it). The intention is that any developer should easily be able to make a 3D model accessible to visually impaired users by using this software. The users would then make use of the 3D Viewer described in the next section to actually access the model.

3. The ENABLED 3DViewer

This program is to be launched by the Enabled architecture but it can also be run as an independent program. The program is implemented making use of the open source api H3D for the haptics and Microsoft Speech - SAPI 5.1 for the speech. It is currently working with the PHANToM Omni, but could be extended for use with the Novint Falcon. The 3DViewer is intended for a severely visually impaired user, and makes use of a keyboard interface together with the two buttons available on a PHANToM Omni. It has the following functions:

- H-key, speech information about the available functions
- Front PHANToM button, say the name of the object with the midpoint closest to the PHANToM pointer (when in move mode – see below – the front button is instead used to drag the model)

Back PHANToM button, drag you to the closest object (again it is the midpoint that is used)
L-key, enlarge. The PHANToM position relative to the object is the same before and after the operation.
S-key, make smaller. The PHANToM position relative to the object is the same before and after the operation.
0-key (zero) to restore the original size and position.
M-key activates/deactivates move mode. In move mode move the model either by click & drag using the front PHANToM button or with the arrow, page up and page down keys
Q-key to change between side view and top view.
A-key to switch audio on or off.
V-key to switch voice (speech) on or off.
Escape key to quit.

These key assignments are as much standardized as possible within the ENABLED system. The study [14] led us to add audio feedback to the zooming which tells you the size by playing a note on the windows synthesizer. To help users know which was the original size a different type of note was played for this size. We also made use of observations from [13] that indicated users felt higher frequencies should correspond to smaller sizes. In the move mode we have added a haptic intertial effect combined with a short “snap” to the PHANToM drag to make users understand better that they are moving the model. When using the arrow keys the user gets speech feedback which tells which kind of move has been done (left, right, in, out, up, down).

The guiding function implemented is very simple compared to what is described in e.g [16]. The reason for this was the wish to keep the interface of the editor as simple as possible. For the attractive force we made use of the force design recommendation from [17] and used a constant attractive force. Figure 2 shows a screen shot showing a map environment from above.

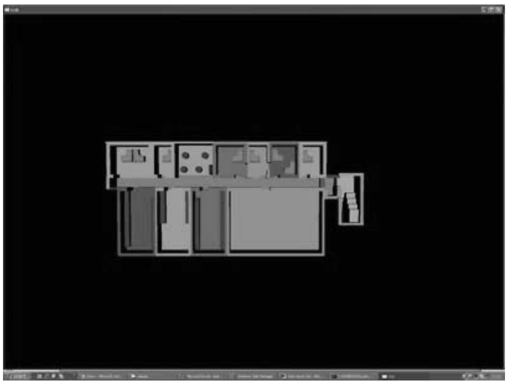


Figure 2. A screen shot from the ENABLED 3DViewer showing a 3D map from above.

4. User testing of the 3DEditor

Due to time constraints it was decided to focus on testing the Viewer, and the Editor test was done with only three developers (Nielsen [18] recommends 3-5 users to screen for serious design flaws). Two developers were male and one female. All had experience from 3D development and VRML/X3D, two were familiar with assistive technologies and one was not, and one had some web developer experience (the other had also done some web work, but did not specify themselves as web developers). All were programmers (two experienced and one less experienced). One had long experience from haptic programming, one had a little and one had no experience at all from this type of programming. Two developers were in the age range 26-40 and one in the range 41-60.

The test persons were given a short introduction about what the program was expected to do. After this they were given the manual and were instructed to make use of it in case of problems. Then the users were asked to perform the following four tasks:

- Open the provided x3d file.
- Adjust the size and position of the model so that it fits well into the workspace.
- Make it possible to touch the model with the PHANToM.
- Add names to 3 objects on the map.

5. Results of the Editor test

All three users were able to complete all the tasks. There were two main problems:

1) The model was initially quite large and far away, and the users did not know initially what it was supposed to look like. Because of this it was not easy to understand the feedback given and it took a little time before they got things right.

2) The users did not look for a command to change the view, and needed a hint about this.

Otherwise the operations are easy to perform and no serious problems occurred.

6. User testing of the 3DViewer

Ten severely visually impaired users (5 in Sweden and 5 in Spain) participated in these tests (the age distribution is summed in table 1). 9 of the 10 were Braille readers. Two were adventitiously blind, five were congenitally blind and three were partially sighted. Four users were female and six were male. All had tested haptic technology before. Four of the five users in Sweden had used the PHANToM quite a lot, and can be considered more experienced PHANToM users. Also in Spain the users had some previous experience of haptics since they had all tested haptic applications before.

16 or under	**17 - 25**	**26 - 40**	**41 - 60**	**61 - 80**	**Over 80**	
1	2	4	3			

Table 1. Age distribution of the test users.

All users except one were used to Braille displays, and 8 had experienced stereo type sounds/3D sounds in the computer environment before. 8 of the users had also tested (or owned) mobile technologies (mostly phones or GPS/navigational devices). Seven users said they had received special training on reading tactile images.

The whole test session was maximized to two hours. During the test one test leader and one observer was present (apart from the test person). Test results were recorded in written form by both the observer and the test leader. Before the test session the user signed the informed consent form.

Due to the users previous experience of the PHANToM and of haptic applications, no special training was given beforehand. A short intro on the program was given where the main functions in the 3D Viewer was explained. In this intro the following functions were demonstrated:

- Front PHANToM button say the name of the object one is in contact with (when in move mode – see below – the front button is instead used to drag the model)
- Back PHANToM button drags you to the closest object (again it is the midpoint that is used)
- Press V-key to switch voice (speech) on or off. Pressing the front PHANToM button will still give "speech on demand" even if the voice is turned off.

Added to this the user was told that it was possible to perform panning and zooming operations, and that these could be demonstrated on demand. The user was also told that he/she could always press the H key to get help from the program (the program also says this by itself when it starts). The first test task was designed to provide some initial training, since it was an open task where the user was asked to explore and describe. The test environment can be seen in figure 2.

The test tasks were:

1. Explore the map. Give an overview description of it.
2. Start in the stairs, enter the lobby and then enter the corridor. Count the number of rooms at the right hand side. How many?

3. Start at the same place as in 2. Count the number of rooms on the left hand side. How many?
4. Find the largest room - which is it?
5. Describe the way from the stairs to the room 204.

7. Results of the Viewer test

All users except two were able to complete the final task. Of these two one was fooled by the speech feedback, while the other was not able to find the room at all. The descriptions generated are summarized in table 2 (the room 204 is the room just to the left of the purple open area with the four round red tables in figure 2).

Test person	**Answer**
P1	Second last on the right hand side
P2	Through the lobby, right side, 5th door just after the open area
P3	Go past the open area with tables and chairs. It is the next room on the right after that.
P4	Go into the corridor and past the open area with tables and chairs. It is the next room on the right after that.
P5	Fifth room on the right, just after the open area.
P6	Corridor after the coffee-room
P7	- (cannot find it)
P8	Follow the corridor straight, right until 4th door (She receives the speech feedback "room 204" on the left wall of the open-area, so she thinks wrongly she is in the room 2004).
P9	After the coffee-room, to the right
P10	By the corridor, just after the coffee-room

Table 2. Description of how to find room 204.

The overview descriptions (task 1) of the environment can be found in table 3. Task 2 presented some problems and six users did not get the number of rooms right. Task 3 presented less problems and all users got that number right. Task 4 was also easy and all users managed to find the largest room.

Test person	**Answer**
P1	A corridor with rooms
P2	A corridor with rooms
P3	Corridor with rooms
P4	A corridor with rooms on both sides
P5	On the left a long room followed by 3 shorter rooms. On the right two rooms and then an open area and after that 3-4 rooms (then you are back where you started)
P6	A large horizontal corridor, 4 doors at a side and a corridor in vertical
P7	Corridor with rooms that have tables
P8	Long Corridor with 7 rooms, open space with several tables
P9	Corridor with doors in both sides, lobby to right and another room without name to the left
P10	Horizontal corridor, rooms to both sides

Table 3. Overall descriptions.

Users were also asked about the best and worst features of the Viewer. The best features were:

⇒ That you get an image in your head and know where the rooms are
⇒ Easy to find the rooms
⇒ That you see it from the top
⇒ That you can hop over the walls - this way you can really check the layout. Good that tables are shown. After a while it is better to turn the speech off (but initially the automatic speech is good)
⇒ Easy to get an idea of the map layout. Good that you get the room names already out in the corridor (you don't have to enter the rooms).
⇒ Audio help with the names
⇒ She likes the final goal of the application that allows her to explore the different levels of an indoor environment, although she considers it can be improved a lot
⇒ Spatial representation (3D). This would be great for railway stations and airports
⇒ 3D Representation

while the worst features were:

⇒ The pen did not always work (the back phantom button had stopped working so the attractive force did not work)
⇒ Don't know

⇒ That the names of the rooms were said out in the corridor - should only say it at the door
⇒ Hard to say - maybe harder walls.
⇒ Hard to say - the pen is hard to hold - would prefer thimble eg.
⇒ Voice synthesis is not good. You go out through the floor
⇒ Voice synthesis is not good.
⇒ Voice synthesis is not good.
⇒ Bad sound quality
⇒ Continuous audio feedback is annoying (useful to be able to deactivate it). Sound quality bad.

It is encouraging that the worst features are things which are not really features of the viewer, but rather things like the voice (which depends on which voice the user has installed in his or her computer), non-working hardware and the fact that you get room names on the outside walls (which depends on how you design your map).

Improvements suggested are:

⇒ Give a verbal overview in the beginning
⇒ Don't know
⇒ Remember to check it does not collide with Jaws (comment: this has been done)
⇒ Maybe one should use textures more to differentiate different objects.
⇒ Don´t go out. Listen the name of the room without need of coming in (he deactivated the automatic audio help and when he requested this information at the wall of a room, at the corridor, he did not receive the name of the room)
⇒ No 3D. 2D is enough to explore the layouts of the different floors
⇒ The voice. Include textures to delimite/differenciate spaces
⇒ The voice. Omni forces are soft. Don´t jump walls.
⇒ Textures in the big rooms, rest room,..

These improvements can all be taken into account by improved design of the 3D model (even the 2D one, since it is possible to do a more 2D-type design if this is requested), or by installing a better voice for the speech feedback.

8. Discussion

The Editor test showed that developers who have had no tutorial and uses the program for the first time are quite able to do the tasks required to prepare a 3D model exported from 3D StudioMax for use with the 3D viewer. All users think the program is good, and would like to use it for doing this type of maps. While there is room for improvement, the important functions “add haptic surfaces” and the ability to annotate objects are very easy and intuitive. A little practice (e.g with the existing tutorial) can be expected to further improve things.

If we look at the Viewer test, the users were in general able to explore and understand the environment. The program was easy to use despite the very short introduction and the absence of training (confirming that it is a program that is easy to use). We had some hardware problems with the second Omni button which unfortunately caused problems during the first testing Sweden (the attractive force that should drag the user to the objects could not be activated because of this).

It was interesting to observe that for the more experienced PHANToM users the absence of a limiting box was not really a problem – even if they lost touch with the object they were usually able to get back to it quickly (usually they just reversed the gesture that made them lose contact). This leads us to believe that a limiting box is mostly a beginner feature – experienced users are not that dependent on it – this is in contrast with the recommendation in [3].

The map design generated some problems – to indicate endpoints in the corridor these doors were designed to be jumped over. This was not a good idea, and this design needs to be improved. The outside walls said the name of the rooms which generated problems for some users while others liked it. Design guidelines needs to be generated to improve the usability of this type of models.

Still, all users are able to get some kind of overview of the environment, and 8 out of ten were able to find the requested room in the final task (9 out of ten if we count the person who was fooled by the speaking walls). This is a clear indication of the usefulness of this type of maps (further confirming the results from [10]).

9. Conclusion

Both the Editor and Viewer are seen to work quite well. It is possible for a developer to quickly start working with the editor.

As could be expected from earlier studies such as [8,10] the Viewer was well received by the users who are able to use it to understand an environment, get an overview and locate a specific room (only one user failed completely on this task). User comments show the importance of the design of the 3D map and care should be taken when designing this type of environments.

10. Acknowledgements

The reported work has been carried out with the financial assistance of the European Commission which co-funded the IP “ENABLED”, FP6 - 2003 - IST - 2, No 004778. The authors are grateful to the EU for the support given in carrying out these activities.

References

[1] C. Ramstein, O. Martial, A. Dufresne, M. Carignan, P. Chasse, and P. Mabilleau, “Touching and Hearing GUI’s: Design Issues for the PC-Access System”, ACM ASSETS 1996, ACM Press, pp. 2-9.

[2] L. Rosenberg, and S. Scott, “Using Force Feedback to Enhance Human Performance in Graphical User Interfaces”, Proc. CHI 1996, ACM Press, pp. 291-292.

[3] Sjöström, Calle, "Non-Visual Haptic Interaction Design - Guidelines and Applications." Ph D Dept. for Design Sciences, Lund University, Faculty of Engineering, 2002.

[4] Yu, W. and Brewster, S. A. Comparing Two Haptic Interfaces for Multimodal Graph Rendering. 2002. Florida, USA.

[5] Yu, W., Kangas, K., and Brewster, S., "Web-based haptic applications for blind people to create virtual graphs," Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2003.HAPTICS 2003.Proceedings.11th Symposium on, pp. 318-325, 2003.

[6] Bussell, L. Touch Tiles: Elementary Geometry Software with a Haptic and Auditory Interface for Visually Impaired Children. 512-515. 2006. 2003.

[7] Iglesias, R., Casado, S., Gutierrez, T., Barbero, J. I., Avizzano, C. A., Marcheschi, S., and Bergamasco, M., "Computer graphics access for blind people through a haptic and audio virtual environment," Haptic, Audio and Visual Environments and Their Applications, 2004.HAVE 2004.Proceedings.The 3rd IEEE International Workshop on, pp. 13-18, 2004.

[8] Magnusson, C., Rassmus-Gröhn, K., Sjöström, C., and Danielsson, H. Navigation and Recognition in Complex Haptic Virtual Environments - Reports from an Extensive Study with Blind Users. Wall, S. A., Riedel, B., Crossan, A., and McGee, M. R. 2002. Edinburgh, UK.

[9] Magnusson, C. and Rassmus-Gröhn, K. Audio haptic tools for navigation in non visual environments. 2005. 11-17-0050.

[10] Magnusson, C. and Rassmus-Gröhn, K., "A Virtual Traffic Environment for People with Visual Impairments," Visual Impairment Research, vol. 7, no. 1, pp. 1-12, 2005.

[11] Simonnet, M., Vieilledent, S., Guinard, J-Y., Tisseau, J., “Can haptic maps contribute to spatial knowledge of blind sailors ?”, ENACTIVE 07, Grenoble, France, November 19-22, 2007.

[12] Kamel, Hesham M., "The Integrated Communication 2 Draw (IC2D)." Ph D Electrical Engineering and Computer Sciences Department, University of California, 2003.

[13] Rassmus-Gröhn, K., Magnusson, C., Eftring, H. “AHEAD - Audio-Haptic Drawing Editor And Explorer for Education”, HAVE 2007 – IEEE International Workshop on Haptic Audio Visual Environments and their Applications, Ottawa – Canada, 12-14 October 2007

[14] Magnusson, C., Gutierrez, T., Rassmus-Gröhn, K., ”Test of pan and zoom tools in visual and non-visual audio haptic environments, ENACTIVE 07, Grenoble, France, November 19-22, 2007.

[15] C.C. Tan, W. Yu, and G. McAllister, “Developing an ENABLED Adaptive Architecture to Enhance Internet Accessibility for Visually Impaired People”, Proc. ICDVRAT 2006, pp. 231-238.

[16] Pokluda, L., Sochor, J., (2005) Spatial Orientation in Buildings Using Models with Haptic Feedback, Proceedings of WorldHaptics 2005 (Pisa, Italy, 2005), pp 523-524

[17] Magnusson, C. and Rassmus-Gröhn, K., Force design for memory aids in haptic environments, ENACTIVE 07, Grenoble, France, November 19-22, 2007.

[18] Nielsen, J, Usability Engineering, 1993, Academic Press.

A Low-Cost Gesture Recognition System for Rehabilitation and Movement Assessment

Paolo Tripicchio* Otniel Portillo-Rodriguez* Oscar O. Sandoval-Gonzalez*
Emanuele Ruffaldi* and Carlo A. Avizzano*
(*) Scuola Superiore Sant'Anna, Italy
E-mail: p.tripicchio@sssup.it, otniel@sssup.it

Abstract

Nowadays there is a plethora of available tracking devices used for gesture recognition, however the most of them are expensive systems. This study was intended to investigate a low cost gesture recognition device that could be used in clinical applications. For such reason, in this paper is proposed an online system based on low-cost accelerometers to capture, assess and recognize simple hand gestures. As application is presented a virtual reality system to be used in rehabilitation of children affected by Traumatic Brain Injury(TBI).

1. Introduction

Gesture recognition is one of the most attractive topics of investigation in computer science. It has strong potential on applications with human machine interfaces (HMI), which interact with the user in natural ways without the need of mechanical devices. Gesture recognition is primarily focused on hand, upper body and face. But actually investigations of whole body action [14] have been performed. Moreover, postures and proxemics can also be topic of gesture recognition. An interesting field of application can be found on movement assessment in clinical settings where traditional approaches are labor intensive and not reliable [6].

Currently systems which make use of markers to capture motion have become very common. They are classified into two categories: (1) Optical systems (passive markers, active markers, time modulated active marker, etc.) and (2) Non-optical systems (inertial, mechanical motion and magnetic). For example in optical system reflective markers are used together with infrared light sources to estimate positions [15]. However these systems require users to wear a large number of markers all over the body in proper patterns. The whole system, although having an expensive price, should be extensively configured before some one can use it. On the other hand, optical markerless motion capture systems interpret the motion of human using one or more cameras that track the subject through the recognition of some features in the images [23]. Markerless motion capture has been widely used for a range of applications in the surveillance, film and game industries. However the use of such systems in clinical assessment is prohibitive due to the cost of commercial top-end devices. Recently new devices that use integrated accelerometers sensors are emerging; such is the case of the commercial Wii remote controller [1]. As is well known the accelerometers sensors are able to detect sudden changes; furthermore, they can be used to calculate velocity and position. However they require high frequencies of sampling to be accurate in a position system, as well as they have to use some techniques to cope with position drift.

The use of virtual reality systems applied to rehabilitation is considerably growing in recent years cause it could presents many possibilities for flexible low cost solutions as well as more complex immersive systems [5]. Patients with TBI often show impaired motor planning when brain networks responsible for motor intensions and translation into movements are damaged. This lead to increased response time and variability, poor trajectory control and so on [13]. The importance of assessing individual differences is fundamental cause the kinematic and functional level deficits may vary from patient to patient, hence, movement skill can be retrained by shaping motor planning functions. The use of VR systems can engage the patient, enhance motivation and automate data collection [3].

In literature many methodologies for motion and gesture recognition are discussed, the most common used comprise template matching [18], feature-based neural networks [17], statistical analysis [12] and hidden Markov models [8, 21, 10]. Recently some approaches related to gesture recognition with the use of accelerometers have been researched. Rehm and colleagues [16] analyzed the patterns of users gesture expressivity to uncover their cultural background. Unzueta at al. [22] proposed a method to distinguish static or quasi-static poses from dynamic hand actions. Barbieri and colleagues [4] created a power motion capture system which consisted of integrated accelerometers and a palmtop computer to track a human's arm motion. Hollar [11] has developed an accelerating sensing glove to demonstrate that accelerometers can be used to translate finger and hand gestures into computer signals.

The presented recognition system is able to recognize gesture starting from acceleration data without the need of precise reference and constraint for the user. This allow to use the system even in complex 3d spacial gestures with the possibility to combine many devices and reconstruct complete full body gestures. All previously presented paper make use of infrared sensor data to calculate a position reference so this limit the application cause the Wii remote controller should be oriented toward the sensor bar. In particular with the implemented system is possible to retrieve quality parameters from the gesture so that is simple to perform assessment for clinicians.

The next sections depict the presented gesture recognition system, there will be an overview on the technology used. Afterward will be discussed the algorithm and finally it will be presented a simple rehabilitation application.

2 Technology

The device used to acquire gestures is the Wii remote controller (wiimote). The wiimote offers several interaction paradigm due to the use of accelerometers and optical sensor technology. This device communicates with the computer via Bluetooth protocol, which facilitates data transmission over short distance for fixed and mobile devices. The Wii remote controller can sense both rotational orientation and translational acceleration along x, y and z axis via a 3-axis accelerometer (ADXL330) embedded in the controller. The accelerometer has a measurement range of +/- 3 g minimum and outputs analog voltage

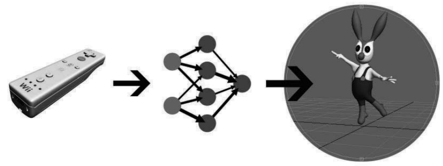


Figure 1: Concept of the Application

signals proportional to acceleration. The controller also features a 1 megapixel sensor with an infrared filter that is commonly used to sense position thanks to an infrared sensor bar installed over the screen. We decided not to use it to allow patients perform gestures without orientation or position constraints.

To communicate with the wiimote via a PC we have chosen the WiiYourself [2] library from the opensource libraries available on the net. Than we have written a demo program to get the raw data from the controller and send the data trough UDP protocol to a Simulink program. The Simulink program is responsible for performing the recognition of the gesture implementing a neural network that will be explained in the next section.

Once recognized the gesture the program send the gesture identifier to a VR application and logs performance variables for evaluation of the quality of performed gestures. The VR application was developed in XVR [7] , such a framework was chosen since its performances were been proven in many Virtual Reality projects running on the Web and in immersive VR installations. The use of UDP protocol for interprocess communications guarantees fast transmission rate that are necessary for real-time interactive applications.

3 Algorithm

In order to assess how a gesture is done by the user and what gesture is done by the user, we study the expressivity of the gesture performed. How a gesture is done, can be described by what Gallaher [9] calls expressivity or expressive style. Gallaher categorizes gestural style by a number of expressivity parameters, e.g. how fast a gesture is done, how much space one uses to perform a gesture, and links expressive style to the quality of the performed gesture.

To obtain the selected expressivity features from the user's gesture in a format we can use, we must transform it from raw data into the following feature variables: power, spatial extent, speed, length of signal, mean of each acceleration axis.

Power, spatial extend and speed were calculated from the S and L variables defined by:

$$S = \sum_{i=0}^{n} a_{x_i}^2 + \sum_{i=0}^{n} a_{y_i}^2 + \sum_{i=0}^{n} a_{z_i}^2$$

$$L = \sum_{i=0}^{n} \|a_{x_i}\| + \sum_{i=0}^{n} \|a_{y_i}\| + \sum_{i=0}^{n} \|a_{z_i}\|$$

Then,

$$Power = \frac{1}{n}S$$

$$SpExt = \frac{S}{L}$$

$$Speed = \frac{1}{n^2}SL$$

Where n is the number of samples of the gesture. The discussion of the how these parameters are useful for the recognition of the gesture and its expressivity are described in detail in [16].

The problem of recognizing the featured variables in a given gesture was solved using Probabilistic Neuronal Networks (PPN). In a PNN when an input is presented, the first layer computes distances from the input vector to the training input vectors and produces a vector whose elements indicate how close the input is to the training input. The second layer sums these contributions for each class of inputs to produce as its net output a vector of probabilities. Finally, a competitive transfer function on the output of the second layer picks the maximum of these probabilities, and produces a 1 for that class and a 0 for the other classes [20].

One of the greatest advantages of this network is that it does not require any iterative training; it trains quickly since the training is done in one pass of each training vector, rather than several. However, one of the main disadvantages of this network is that it has one hidden node for each training instance and thus requires more computational resources (storage and time) during execution than other models.

PNN can be used in real time because as soon as one input vector representing each class has been observed, the network can begin to generalize to new input vectors. As additional input vectors are observed and stored in the net, the generalization will improve and the decision boundary can get more complex [19]. In order to test the performance of the PNN to recognize gestures; two different data sets were obtained, one is for training and the other is for test. Each data set is formed by all the feature vectors extracted of 20 trials for each gesture performed.

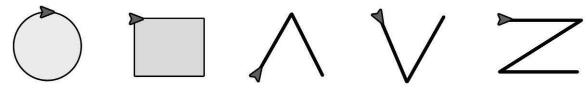


Figure 2: The set of recognized gestures

Using the data training set and its corresponding class gestures 100 instances of the one PNN were created, the architecture of this neural network is fixed, and depends only of the number of elements in the input vector, the number of training feature vector and the number of gestures to identify. Then the data test set was feed to all the PNN instances and provided always the same results, although the creation and evaluation times were different. Table I shows the average results obtained for 100 instances of the PNN network. In general this network is fast to be created and trained obtained a 95% of accuracy. All of the tests were performed in Matlab version 7.3 on a PC with a 1.8 Ghz Centrino microprocessor and one gigabyte of memory.

Time evaluation of the test matrix	Time Creating and Training	Percent Error
0.0142 s	0.0398 s	4.3943

Once the PNN was identified useful as gesture recognizer, an online application developed in Simulink was done. In this application the user can capture gestures either to train or test the PNN. When the user perform the gesture, all the expressivity parameters are calculated on line, once the gesture is finished, the feature parameters are sent to the PNN which responds the gesture that is the most similar to the performed movement.

4 Application

With the presented gesture recognition system is possible to develop many VR interactive application that make use of the gestures to operate in the virtual environment, however our purpose was to develop a system intended for rehabilitation of child TBI patients. In particular choosing simple gestures that could activate specific brain networks and specific muscles to be trained. The gestures chosen for the demonstrator are represented in figure 2 where the red arrow represent the starting point of the intended gesture to be performed. Working with children is necessary to engage them and

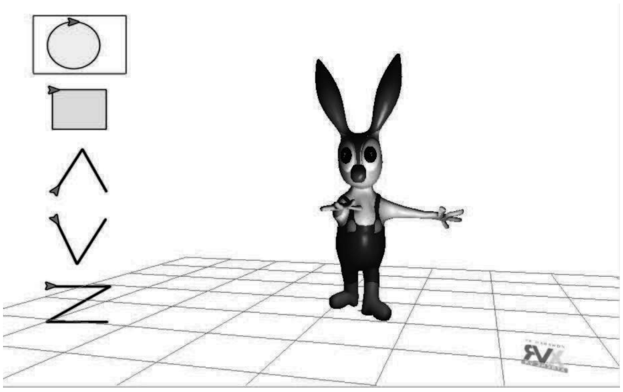


Figure 3: The VR Application

enhance their motivation to focus their attention on the task to complete and not to bore or stress them. For this reason we decided to adopt into the application a funny avatar to mask the training experiments as a simple game. In particular the avatar responds to the gesture performed with a specific animation for each gesture so that if the gesture is not well performed the child is motivated to repeat it to see the reaction of the avatar.

Every animation was obtained trough motion capture data so that the movement of the avatar is realistic, for example when the application recognize that the patient have performed a circular gesture the avatar start to dance doing some pirouettes (Figure 3).

5 Conclusions

The paper presented a low-cost gesture recognition system that is able to recognize gesture from accelerometers data. This system was used in a VR rehabilitation application suitable for children with an interaction device capable of acquiring user's motion, the Wii remote controller. The developed neural network was then capable to recognize and assess correct gestures. Future work will investigate gesture recognition of full body movements trough the use of multiple controllers. From the application point of view it will be possible to focus the attention on clinical patterns involving therapists during the developing of the application.

6 Acknowledgments

We wants to thanks the ENACTIVE Network of Excellence European Project and the EU Integrated Project SKILLS, that deals with the use of MultiModal systems for the capture model and rendering of data signals for the transfer of skills in different application fields.

References

[1] http://en.wikipedia.org/wiki/wiimote.

[2] http://wiiyourself.gl.tter.org/.

[3] K. K. C. M. Albert Rizzo, Maria Schultheis. *Neuropsychological Rehabilitation*, volume 14, chapter Analysis of assets for virtual reality applications in neuropsychology, pages 207–239. Psychology Press, part of the Taylor & Francis Group, 2004.

[4] R. Barbieri, E. Farella, L. Benini, B. Ricco, and A. Acquaviva. A low-power motion capture system with integrated accelerometers [gesture recognition applications]. *Consumer Communications and Networking Conference, 2004. CCNC 2004. First IEEE*, pages 418–423, Jan. 2004.

[5] S. D. W. S. L. Blanton S., Porter L. *Rehabilitation of the adult and child with traumatic brain injury*, chapter Strategies to enhance mobility in traumatic brain inured patients, pages 219–241. Philadelphia: F. A. Davis Company, 1999.

[6] B. Bobath. *adult hemiplegia: evaluation and treatment*. London: Heinemann Medical Books, 1989.

[7] M. Carrozzino, F. Tecchia, S. Bacinelli, C. Cappelletti, and M. Bergamasco. Lowering the development time of multimodal interactive application: the real-life experience of the xvr project. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology*, pages 270–273. ACM Press New York, NY, USA, 2005.

[8] H. Deng, J.W.; Tsui. An hmm-based approach for gesture segmentation and recognition. In *ICPR '00: Proceedings of the International Conference on Pattern Recognition*, page 3683, Washington, DC, USA, 2000. IEEE Computer Society.

[9] P. E. Gallaher. Individual differences in nonverbal behavior: Dimensions of style. *Journal of Personality and Social Psychology*, 63(1):133145, 1992.

[10] Y. Iwai, H. Shimizu, and M. Yachida. Real-time context-based gesture recognition using hmm and automaton. *Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 1999. Proceedings. International Workshop on*, pages 127–134, 1999.

[11] S. H. K. P. J. Perng, B. Fisher. Wireless static hand gesture recognition with accelerometers the acceleration sensing glove. In *ISWC International Symposium on Wearable Computers*, San Francisco 1999.

[12] J. Martin, D. Hall, and J. L. Crowley. Statistical gesture recognition through modelling of parameter trajectories. In *GW '99: Proceedings of the International Gesture Workshop on Gesture-Based Communication in Human-Computer Interaction*, pages 129–140, London, UK, 1999. Springer-Verlag.

[13] H. A. J. McCrea P. H., Eng J. J. *Disability and Rehabilitation*, volume 24, chapter Biomechanics of reaching: clinical implications for individuals with acquired brain injury, pages 534–541. Informa Healthcare, 2002.

[14] U. L. S. B. Mena, O. and L. Matey. Temporal nearest end-effectors for real-time full-body human actions recognition. In *AMDO 2008 V Conference on Articulated Motion and Deformable Objects. Andratx, Mallorca, Spain*, July 9-11, 2008.

[15] H. Park and J.-I. Park. Invisible marker tracking for ar. In *ISMAR '04: Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 272–273, Washington, DC, USA, 2004. IEEE Computer Society.

[16] M. Rehm, N. Bee, and E. Andr. Wave like an egyptian accelerometer based gesture recognition for culture specific interactions. In *Procedings of HCI 2008 Culture, Creativity, Interaction*, in press.

[17] D. Rubine. Specifying gestures by example. *SIGGRAPH Comput. Graph.*, 25(4):329–337, 1991.

[18] G. S. Schmidt and D. H. House. *Gesture-Based Communication in Human-Computer Interaction*, volume 2915/2004 of *Lecture Notes in Computer Science*, chapter Model-Based Motion Filtering for Improving Arm Gesture Recognition Performance, pages 210–230. Springer Berlin / Heidelberg, 2004.

[19] D. F. Specht. Probabilistic neural networks for classification, mapping, or associative memory. *IEEE International conference on Neural Networks*, 1:525–532, 1988.

[20] D. F. Specht. Probabilistic neural networks. *Neural Netw.*, 3(1):109–118, 1990.

[21] G. R. Stefan Eickeler, Andreas Kosmala. Hidden markov model based continuous online gesture recognition. In *ICPR '98: Proceedings of the 14th International Conference on Pattern Recognition-Volume 2*, page 1206, Washington, DC, USA, 1998. IEEE Computer Society.

[22] M. O. S. B. Unzueta, L. and A. Suescuen. Kinetic pseudo-energy history for human dynamic gestures recognition. In *AMDO 2008 V Conference on Articulated Motion and Deformable Objects. Andratx, Mallorca, Spain.*, 2008.

[23] A. Wilson and A. Bobick. Using configuration states for the representation and recognition of gesture. In *Technical Report No. 308, M.I.T. Media Laboratory, Cambridge, MA.*, 1995.

Spatial Sound and Frequency Distortion for Enactive Transfer of Skills in Tai-Chi Movements to Visual Impaired People

Oscar Sandoval-González, Otniel Portillo-Rodriguez, Luis Montes-de-Oca, Paolo Tripicchio, Emanuele Ruffaldi, Carlo A. Avizzano and Massimo Bergamasco

PERCRO, Scuola Superiore Sant'Anna

E-mail: o.sandovalgonzalez@sssup.it, otniel@sssup.it, carlo@sssup.it

Abstract

A novel methodology for Enactive Skills Transmission of Tai Chi movements through sound feed-back stimuli is presented in this paper. We adopted the paradigm of spatial sound and frequency distortion as an accurate methodology for the transmission of error-feedback to the user. An intelligent system based on Gesture recognition methodologies and a real-time motion descriptor system is applied to engage and guide the users to perform an unknown movement in a natural and transparent process using the sound like the base of the transfer of the skill.

Keywords. Spatial Sound, 3D sound feed-back, 3D Gesture Recognition, real-time descriptor, skills transfer using sound.

1. Introduction

Human cognitive processes frequently involve all sensory modalities. Nowadays, multimodal systems, based on the idea of Human-Machine Interaction HMI (1)(2)(3), have been developed in order to augment perception and facilitate cognitive processing and skills transfer. From the cognitive point of view, the imitation process is one of the most effective methods for the machine-human transfer of a skill.

With visual impaired people, the senses of hearing and touch can be used for the purposes of imitation-based learning. Therefore, several studies investigated how to deliver multimodal information to visual impaired people, focusing on haptics. One example of touch and audio haptic device was designed by Iglesias and Casado (4) where visually impaired people had access to the three-dimensional computer graphics displays through the sense of touch, augmented by audio output and voice commands. Fritz and Barner (5) designed a Haptic data Visualization system for people with visual impairments. Similar experiments were carried out by Fernandez and Muñoz (6) who developed several haptics applications for blind people.

Sound is an important component of cognition in the context of human action. Therefore one motivation of this paper is to design a system capable to engage the users in order to perform movements following the harmony of the sound.

Important pilot studies related to spatial structures and virtual worlds through audio interfaces have solved different questions regarding to the mental cognitive models of the human being: Lumbreras and Sanchez (7) developed an application to navigate and interact in a virtual world using spatial sound interfaces. The results have indicated the possibility to use spatial sound to stimulate blind learners to construct navigable spaces in an interactive virtual world. An interesting conclusion of this study is that spatial mental images could be reconstructed with spatial sound and without visual cues. Similar experiments performed by Mareu and Kazman (8) found that the use of 3D audio interfaces by a blind person could help to localize a certain point in the 3D space. In the same field of research Baldis (9) conducted an experiment to determine the effect of spatial audio on memory and comprehension.

Loomis, Lippa, Klatzky and Golledge (10) designed a study with blind and blindfolded sighted observers presenting them auditory stimuli specifying target location by using 3D sound. They assessed the ability to mentally keep track of the target location without concurrent perceptual information about it.

All these researches have demonstrated the cognitive relevance of spatial sound. For experimental purposes the position of the sound in the 3D space and the distortion of the sound through the variation of the frequency (Pitch) are variables that can be easily modified. This paper presents a novel system based on Machine Learning (ML) methodologies which transfer in a common and natural link and in an enactive way, the skill of tai-chi movements to impaired people only using spatial sound. The cores of this system are the Gesture Recognition System (GRS) and the Descriptor Process (DP) which make possible to know what the user is doing, pretending to do in order to identify the necessary methodologies to motivate and teach the students to perform an unknown-movement in a transparent way.

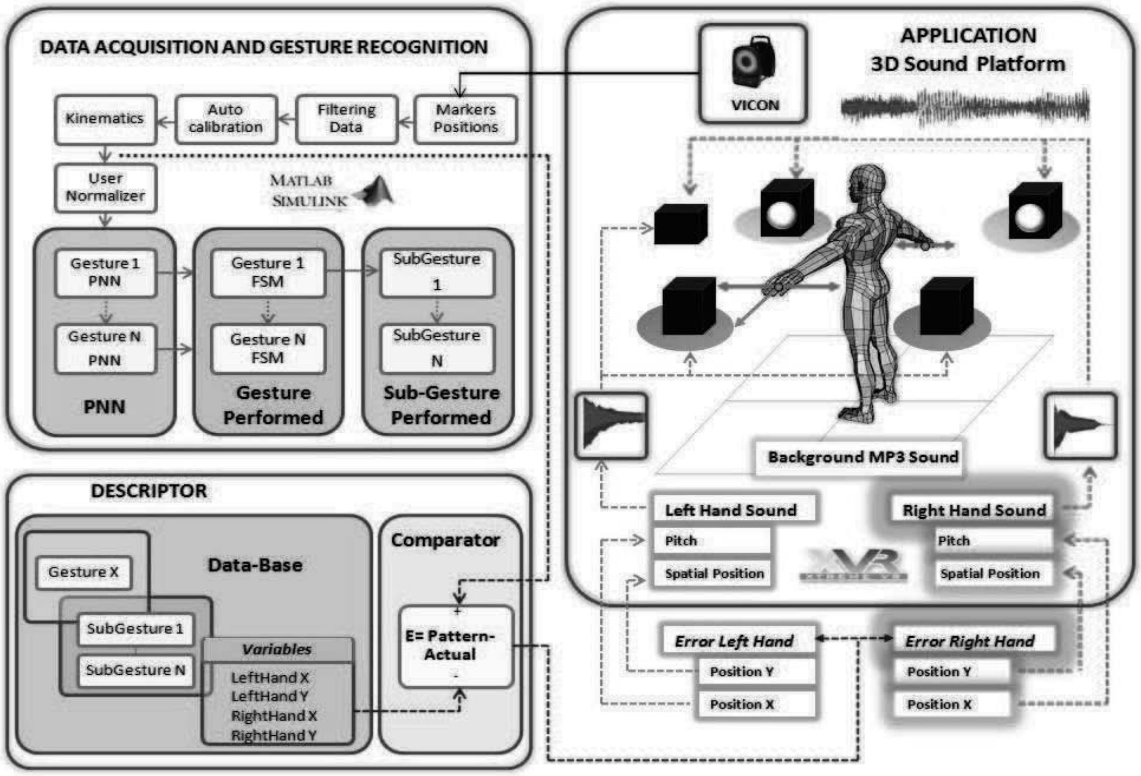


Figure 1. Architecture of the 3D Sound Platform

2. System Implementation

In general terms the gesture recognition system is capable to divide the movement on different states (Time-independent). These steps are evaluated by the descriptor system and audio feed-back stimuli is played in order to correct the user's movements. The Platform is composed by: The hardware and software of the 3D tracking optical system (VICON), the gesture recognition system and the description of motion (both running in Matlab Simulink), and the 3D sound system developed by the XVR 3D sound toolbox.

2.1 Data Acquisition

An optical tracking device (VICON) was used to obtain the spatial position of the markers in the user's body. The VICON system provides millimetric accuracy at 300Hz of data sampling in the 3D space through the use of passive reflective markers attached to the body. The system computes the kinematics model of the body composed by fourteenth DOFs. Normally, it is common to spend a great deal of time creating or modifying the kinematics structure in Vicon according to body size of each user. Therefore, an auto calibration algorithm was designed to reduce the problem. This algorithm obtains the initial position of the markers of a person placed in a military position called "stand at attention".

The algorithm checks the dimension of his/her arms and the position of all the markers. The angles and distances between the joints are computed and finally this information is compared with to the ideal values to compensate and normalize the whole system.

2.2 Real-Time Gesture Recognition Process

In order to recognize the gesture performed by the user, a state space model approach was selected (11)(12). A common issue with the modeling of gesture with this approach is the characterization of the optimal number of states and the establishment of their boundaries. A dynamic *k*-means clustering on the training data defines the number of states and their spatial parameters of the gesture without temporal information (13). This information from the segmented data is then added to the states and finally the spatial information is updated. This produces the state sequence that represents the gesture. The analysis and recognition of this sequence is performed using a simple Finite State Machine (FSM), instead of use complex transitions conditions which depend only of the correct sequence of states for the gesture to be recognized and eventually of time restrictions.

The novel idea is to use for each gesture a PNN to evaluate which is the nearest state (centroid in the configuration state) to the current input vector that represents the user's body position. The input layer has the same number of neurons as the input vector and the second layer has the same quantity of hidden neurons as states have the gesture. In this architecture, each class node is connected just to one hidden neuron and the number of states (where the gesture is described) defines the quantity of class nodes. Finally, in the last layer, the class (state) with the highest summed activation is computed. A complete reference of the whole system is found in (14). A number of sixteen variables were used in our configuration space: 2 distances between hands and 2 between elbows, 2 vectors created from the XYZ position from the hands to the chest and 2 vectors created from the XYZ positions from the elbows to the chest.

2.3 Real-Time Descriptor Process

The comparison and qualification in real-time of the movements performed by the user is computed by the descriptor system. In other words, the descriptor analyzes the differences between the movements executed by the expert and the movement executed by the student, obtaining the error of the values and generating the feedback stimuli to correct the movement of the student.

$$\theta error = [P(n+1) - U(n) * Fn] \quad (1)$$

Where *θerror* is the difference between the pattern and the user, *P* is the pattern value, *U* is the user value, *Fu* is the normalize factor and n is the actual state.

The DP can analyze step-by-step the movement of the user and creates a comparison between the movements performed by the master and user. This DP can compute the comparison of 26 variables (Angles (12), positions (12), distance between hands (1) and elbows(1)). For this application we only use 4 measurements which represent the X-Y deviation of each hand with respect to the center of the body. Figure 3 shows the 5 Tai-Chi movements which are recognized by the gesture recognition system and analyzed by the descriptor.

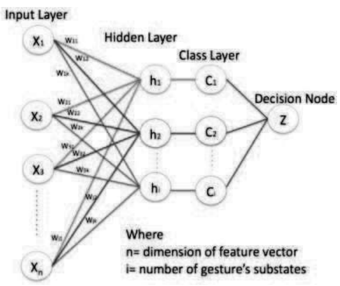


Figure 2. PNN architecture used to estimate the most similar gesture's state from the current user's body position

2.4 Sound Feed-Back Methodologies

We designed an optimized sound methodology which consists in playing in a background plane a relaxed song, "New Age" style, with a slow tempo and non drastic sound variations. In the same plane 2 different sounds, which represent the right and left arm, are located in the corresponding pair of right/left speakers. Each sound has a characteristic rhythm and represents the chords to the background song. Therefore, the sounds remain all the time in harmony when the user is performing a correct movement. However, when the movement is incorrect in the X-Y plane two methodologies (explained in section 2.4.1 and 2.4.2) are applied.

Figure 3. Tai-Chi Movements

2.4.1 Frequency Distortion

For the X deviation we implemented a methodology based on frequency distortion in the following way:

1. The fundamental frequency of the sound of the right and left arm are obtained. These frequencies represent the neutral point where the distortion of the sound will be applied in higher and lower frequencies in order to simulate the perception of error in the movement.
2. Error deviation: The descriptor system computes the error between the Pattern and the user movement of the X-plane in both hands, and obtains the difference.
3. The frequency distortion is applied. This deformation is proportional to the error that was measured by the descriptor system. In other words, in the X-plane, the user will not hear the frequency distortion if the movement performed is correct. However, if the movement has a positive error, the pitch will increase the frequency in proportional steps of 700Hz/cm (700Hz per centimeter of error) and will decrease in proportional steps of 700Hz/cm when the error is negative. This frequency was chosen according to different experiments and surveys that were performed based on the perception of the users, section 3.4 shows in detail the results of these experiments.

$$PX_{LH-RH} = \lfloor(\theta_{px} - \theta_{ax}) * fmc\rfloor + Fc \quad (2)$$

Where PX_{LH-RH} is the Pitch in the X-plane of the Right and Left hand, *θpx* is the Pattern Value in the X-Plane, *θax* is the Actual value performed by the user in the X-Plane, *fmc* is the Minimum Cognitive Frequency and *Fc* is the characteristic frequency of the sound.

2.4.2 Spatial Position of the Sound

For the Y deviation a methodology based on spatial position of the sound was implemented. The idea is to locate the source of sound in a specific position and vary the volume. The strategy consists in the following steps:

1. The positions of the speakers in space are obtained in a manual measuring process.
2. It is determined the position of the user respect to the origin through the Vicon.
3. Sound Calibration: Once the positions of the speakers and the user are known, the volume of each speaker is balanced according to the position of the listener, Eq. 2 shows the calculus.
4. Error Deviation: The descriptor system computes the error between the Pattern and the user movement in the Y-Plane of both hands, and obtains the results.
5. Sound Position. As it shown in Figure 4, when the descriptor obtains a positive value of the error, the algorithm moves the balance of the right/left speakers in a proportional form.

$$Cp_{LH/RH} = Body - \left[\frac{Spk_{1/3} + Spk_{2/4}}{2}\right] \quad (3)$$

Where Cp_{LH-RH} is XYZ coordinates of the Calibration Sound Process for the Left and Right Hand. Body represent the XYZ coordinates where is located the Body, Spk1 and Spk4 are the Rear Left-Right Speakers, Spk2 and Spk3 are the Front Left-Right Speakers.

$$Sp = [Cpx, Cpy, Cpz] + [0, (\theta py - \theta ay) * f_{int}, 0] \quad (4)$$

Where Sp is the Sound Position, Cp is the calibration sound process coordinates, θpy is the pattern value in the Y-Plane, θay is the actual value in the Y-Plane and f_{int} is the Intensity factors which normalize the output.

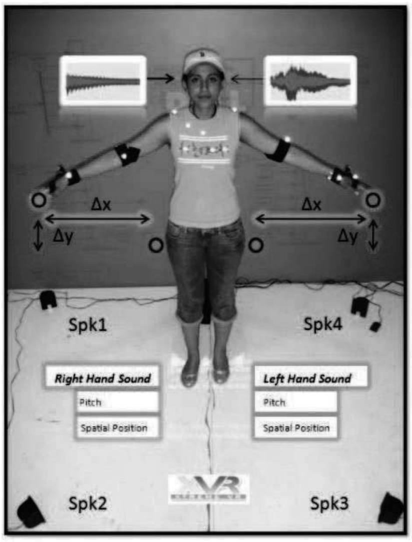


Figure 4. Audio feed-back stimuli when a user is performing a bad movement in the X and Y plane.

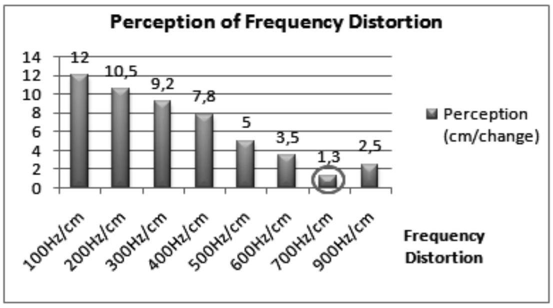


Figure 5. Perception of frequency distortion

3. Experimental Results

We tested our system in a 4-stages experiment. Ten adults without knowledge of tai-chi movements took part to the experiment (9 males, 1 female; age=21-26 years, without sight problems). Across all stages, we assessed to know the perception of the error using the distortion of frequency, the perception of the error using the positional sound, the analysis of the trajectories performed by the user when they tried to find the correct gesture. We finally conducted a survey explained in section 3.4 where the users explain different perceptions that they felt during the experiment.

3.1 Perception of the error using distortion of frequency

The aim in this part of the experiments was to know in detail the perception produced during the feed-back of sound in the distortion of frequency when an error is produced. It is known that many people, especially those who have been musically trained, are capable of detecting a difference in frequency between two separate sounds which is as little as 2 Hz. However, in this experiment we are not interested in knowing the human capability to detect the minimum change of frequency. The fundamental part of the experiment is to determine the cognitive correspondence between the physical motion (the movement of the hands) and the change of frequency produced by the error in the position of the hands. In order to obtain these results, the system establishes one reference point in the actual position of the hands of the user.

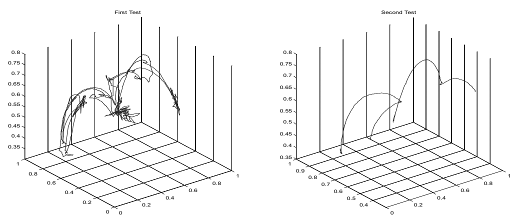


Figure 6. a) Initial trial performed by the user and b) Second Trial

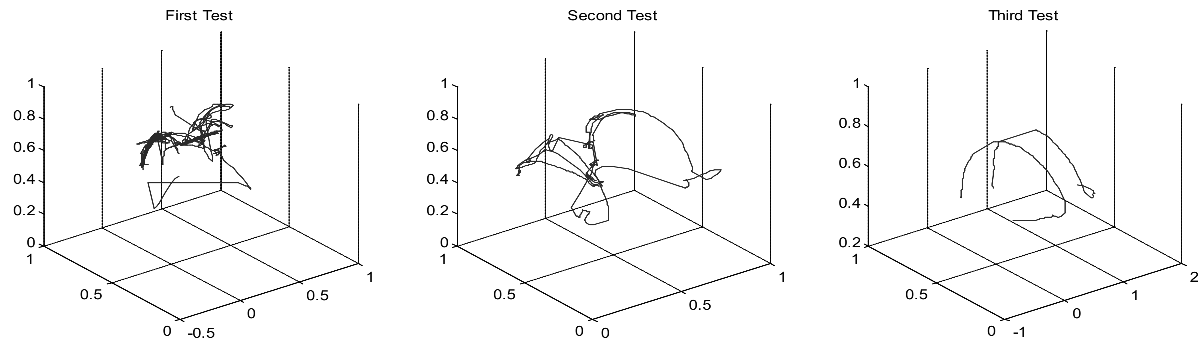


Figure 7 a) Initial trial performed by the user, b) Second Trial, c) Third Trial where the plot indicate an accurate movement.

After this, the users begin to move his/her hands in the X-Plane. The descriptor process recognizes that the hands were located in an incorrect position and computes the corresponding error and the frequency distortion of the sound. At this point when the user moved the hand out of the correct position, he/she feels a distortion of the sound proportionally to the displacement of the hand. Figure 5 shows that the people feel small changes in the correlation of frequency and hand motion in 700Hz/cm.

3.2 Perception of the error using the positional sound

The corrections in the Y-plane are determined by the spatial position of the sound. Normally, when a hand is placed in a forward position with respect to the ideal position in the gesture, the system according to Eq. 4 places the sound in a specific location in the space, that is, in a proportional way increasing the volume of the frontal right/left speaker and reducing the volume of the rear right/left speaker.

In the similar form, when the hand is backward to the ideal position, the system places the sound in the inverse mode. The survey presented in section 3.3 shows that it is difficult to hear small changes in spatial position of the sound only using 4 speakers. However, the perception of error in the Y-plane is detected with this methodology. The accuracy in the perception of the spatial position can be augmented adding more speakers to the system.

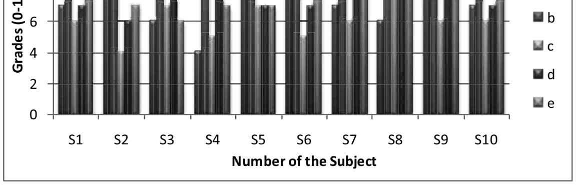


Figure 8. User's perception survey

3.3 Analysis of trajectories

Once the best values related to pitch and spatial position were determined, the next experiment consisted into allowing the user to move alone in order to learn a gesture. This gesture was chosen randomly by the system (the system contains 5 tai-movements), only using these two variables of sound like a feed-back to discover the movement.

As it was expected, the users started to move one of his/her hands in a random way and began to search the position where the error is almost 0, when they arrived to one of the states of the gesture, the system plays a sound. In the same way, the users began to move the hands very slowly in a random direction, all the time trying to compensate the movement through the feed-back stimuli of the sound. For each state of the gesture that they reach, a sound was played and finally when the user reach the final state, the system plays another sound indicating that the gesture have been completed successfully.

When the gesture is completed, in a natural learning process the user is mapping the movement in his/her brain. When the same movement is repeated, the total error in the performance of the movement is reduced a minimum levels. At this moment, the system has transferred the movement to the user using only sound stimuli.

3.4 User's perception survey

The final experiment was performed by a short survey that consisted of 5 questions answered by the 10 users with a score (0-10), where 0 is the minimum and 10 the maximum value. Figure 8 shows the results. The questions were:

a) Perception using Pitch and Spatial Position
b) Perception using only the Pitch, X-Plane.

c) Perception using only the Spatial Position, Y-Plane.
d) Level of identification of the sounds for the Right Hand and Left Hand.
e) Functionability of the system.

3.5 Data Analysis

Figure 9 shows the error analysis (Normalized from 0 to 1) of 5 gestures performed in 5 trials for 10 people. In the first trial the average error of 5 gestures was 0.694 and in the third trial was 0.114.

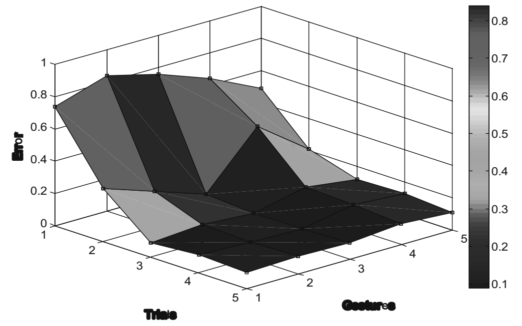


Figure 9. Errors presented by 10 users performing 5 gestures in 5 trials.

The total error in each trial was obtained following the equation (5).

$$ET = \frac{1}{Gs}\sum_{1}^{Gs}\frac{1}{U}\sum_{1}^{U}\frac{1}{Sgs}\sum_{i=1}^{Sgs} abs(Pt[i] - User[i]) \quad (5)$$

Where ET is the Error per Trial, Gs is the total number of gestures, U is the total number of users, Sgs is the total number of Subgestures, Pt[i] is the value of the Pattern in each subgesture and User[i] is the value of the User in each subgesture.

4. Conclusions & Future Work

The results of the experiments conclude that this methodology is an interesting approach in the possible transfer the knowledge of movements from an intelligent 3D sound platform to the human being using an audio feed-back stimulus (pitch and spatial sound variations) according to the methodologies presented in this paper. Moreover, the results support the fact that the system not only helps to the user to identify an unknown gesture but also trains the people through the repetition process. Finally, in Figure 6 and Figure 7 is possible to appreciate notable improvements in the performance of the movements only with few repetitions. Figure 8 shows the statistical decrease of error due to the trials.

For the future work, it is important to perform an appropriate analysis in order to determine in a better form the cognitive process where is involved the audio-motion-action. Another point is that the system must be tested by visual impaired people, in order to analyze their audio cognitive process.

5. References

1. *A System for Medical Consultation and Education Using Multimodal Human/Machine Communication.* Akay, Metin, et al. : IEEE Transactions on information technology in Biomedicine, December 1998, Vol. 2. 1089-7771/98.
2. *Multimodality.* Cole, R. and Mariani, J. : Survey of the State of the Art of Human Language Technolgy.
3. *Toward Multimodal Human-Computer Interface.* Sharma, Rajeev, Pavlovic, Vladimir and Huang, Thomas.: Proceeding of the IEEE, May 1998, Vol. 86.
4. *Computer Graphics Access for Blind People through a Haptic and Audio Virtual Environment.* Iglesias, R., et al.: Haptic, Audio and Visual Environments and their Applications (HAVE) , 2004. 13-18.
5. *Design of a Haptic Data Visualization System for People with Visual Impairments.* Fritz, JP and Barner, K. E.: IEEE Transactions on Rehabilitation Engineering, 1999. Vol. 7. 3.
6. *Haptic Virtual Reality Applications for blind people.* Fernández, J.L., et al.: Revista Integración, 1999, Vol. 29.
7. *3D aural interactive hyperstories for blind children.* Lumbreras and Sánchez. 1,: International Journal of Virtual Reality, Vol. 4. 20-28.
8. Mereu, S. and Kazman, R.: *Audio Enhanced 3d Interfaces for visual impaired users.* Proceedings of CHI, ACM Press, 1996.
9. Baldis, J. *Effects of spatial audio on memory, comprehension, and preference during desktop conferences.* : Proceeding of the ACM CHI , 2001, Vol. 3. 166-173.
10. Loomis, J., et al.*Spatial updating of lacations specified by 3-D sound and spatial language.* 28, s.l. : Journal of Experimental Psychology: Learning Memory and Cognition, Vol. 2. 335-345.
11. Bobick and Wilson, A.F. *A state-based approach to the representation and recognition of gesture.* 12, s.l. : Pattern Analysis and Machine Intelligence, IEEE Transactions, Dec. 1997, Vol. 19.
12. Farmer, J., et al. *State-space reconstruction in the presence of noise*: Physics D. 51D:52 98, 1991.
13. Jain, A.K., Murty, M.N. and Flynn, P.J. *Data Clustering: A review.* : ACM Computing Surveys, 1999, Vol. 31.
14. Portillo-Rodriguez, Otniel, et al. *Development of a 3D Real Time Gesture Recognition Methodology for Virtual Environment Control* : ROMAN, 2008.

Human Perception

Efficiency and Stability of Walking in (Non-) Natural Virtual Reality

Guérin Perrine[*] Bardy Benoit[*†]
(*) Motor Efficiency and Deficiency Laboratory, University Montpellier1, France
Perception and Motor Control Laboratory, University of Paris XI, France
(†) Motor Efficiency and Deficiency Laboratory, University Montpellier1, France*
E-mail: perrine.guerin@free.fr, benoit.bardy@univ-montp1.fr

Abstract

In two experiments, we evaluated the inter-relation between information, energy and locomotion stability during walking in VR. In Experiment 1, we studied the influence of optical flow on the kinematic parameters and the energy expenditure of a fast speed walking. We tested three optical flow conditions: static flow (Static), natural flow, i.e. opposite to locomotion (Natural) and no-natural flow, i.e. reverse to locomotion (N-Natural) into three experimental sessions. Locomotion performed in the Natural condition appeared more efficient than in the other conditions, but this energetic effect decreased over time, indicating locomotor adaptations. Experiment 2 reproduced these results in four new sessions. The efficiency effect of the optical flow conditions (Natural vs. N-Natural) was confirmed, indicating that walking in VR encourages the emergence of the most efficient behavior, in relation with kinematic adaptation.

1. Introduction

In a perception-action coupling perspective, virtual environment technologies are often used to evidence the influence of visual information on human locomotion patterns (e.g., [1], [2]). For example, walking is influenced by optical flow velocity: with large (visual) velocities, human participants tend to decelerate their preferred walking speed [3], modulate stride length or stride frequency [4], increase the variability of the stride cycle [5], transit to the running pattern earlier [6].

The *laws of control* suggested by Gibson [7] and formalized by Warren [8] perfectly capture this interrelation between forces and information: Forward forces applied onto the ground surface create an optical flow at the observation point which contains relevant structure (e.g., expansion, motion parallax) that can in turn be used to modulate the free parameters of the action system. These forward forces need physiological resources and they have an energetic cost.

2. The present study

In a recent study [9], we showed the existence of a significant influence of optical flow velocity on locomotor parameters; in addition, energetic changes were found between visual conditions during a locomotion task performed at preferred transition speed, i.e. during unstable locomotion.

In order to complement the results obtained in [9], we investigated here the specific relation between information (visually detected), the stability of locomotion (expressed by stride length and relative phase variability), and the efficiency of locomotion (energy consumption). Again, because the internal forces applied have optical as well energetic consequences, we expect the manipulation of the coupling between forces and optical flow to have consequences in pattern stability as well as in energy consumption.

The purpose of the present study, composed of two experiments, was thus (i) to evaluate the role of optical flow in stabilizing kinematic and energetic parameters during fast walking and (ii) to evaluate the dynamics of this influence over time.

3. General Method

3.1 Apparatus

The experiments took place in a dark room. Participants locomoted on a motor-driven treadmill at various speeds, facing a projection screen (2.70m horizontal x 1.95m vertical). Raster displays simulating a virtual corridor (see Fig. 1) were generated on a PC workstation, and projected on the screen with a video projector. The subject's field of view was restricted to the screen by a head-mounted mask, which occluded the edges of the display. The virtual corridor was created with the 3D software, and contained familiar objects of known size (e.g. doors, windows) in order to increase the feeling of immersion. The corridor was animated with different velocities and started again at its initial location when reaching its end.

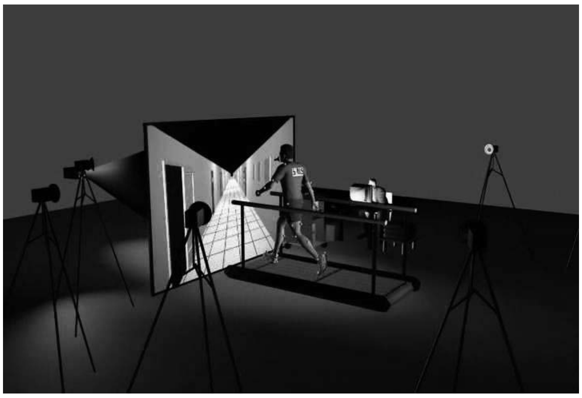


Fig. 1. The experimental device

Three experimental conditions were tested in the two experiments: a static or no flow condition (*Static*) and two conditions in which the direction of projection was manipulated: a natural condition in which the direction of optical flow was opposite to locomotion (*Natural*) and a non-natural condition in optical flow and locomotion were in opposite directions (*N-Natural*). For these two dynamic visual conditions, the velocity of optical flow was constant and determined on the basis of each individual preferred transition speed (PTS) value. The individual optical flow velocity was set up at three times the locomotor speed in order to increase the velocity gradient between locomotor and visual components (e.g., [9]).

In the two experiments, participants were instructed to walk at 110% of their PTS while looking at the far end of the corridor, which served roughly as a fixation point.

3.2 Data acquisition and analysis

For the recording of the kinematic variables, participants had Vicon™ markers pasted on their body. Markers were located on anatomic keys positions: the shoulder (acromion), hip (greater trochanter), knee (estimated knee joint center), ankle (lateral malleolus) and toe (head of fifth metatarsal) of the human body allowing the computation of the main parameters of the locomotor system. Marker positions were recorded in three dimensions with a VICON motion capture system (Oxford Metrics Group). Dependent variables were calculated in each visual condition and included: ankle-knee relative phase and its stability, stride length and its variability. The ankle-knee relative phase was computed using the point-estimate value of relative phase [10].

Concerning the energetic variables, the cycle-to-cycle oxygen data were collected using a ZAN 680 system. During each experimental trial, energy expenditure (*EE*) was calculated relative to the oxygen value scaled to each subject's weight and averaged ($\dot{V}O_2 rest$) during the rest period, using the following equation [11]:

$$EE = (\dot{V}O_2 - \dot{V}O_2 rest).speed^{-1}$$

The average (per cycle) energy expenditure was also computed in each condition.

All variables were computed over the entire trial duration (6 min) but we choose to analyze only the last three minutes of each trial in order to isolate the initial "mise-en-route" (3 min) from the stable phase (3 min) both in terms of kinematics variables and energetic variables.

Repeated-measures analyses of variance were conducted on all dependent variables. We used standard linear and not circular statistics to analyze relative phase values because relative phase variance did not exceed 180°.

4. Experiment 1

4.1 Procedure

Eight adults with normal or corrected vision, and no motor disorders, participated in this experiment.

This first experiment consisted of three experimental sessions. In Session 1, participants were familiarized with the experimental task of walking on the treadmill looking with various optical flow velocities. During this session, their preferred transition speed was determined by an incremental protocol [12]. Sessions 2 and 3 were identical and consisted of three trials of six minutes. In each session, one trial corresponded to one optical flow condition: *Static, Natural or N-Natural*. The order of trials was randomized within each session. For each trial, participants were constrained to walk at fast speed (110% PTS), separated by four minutes of rest.

4.2 Results

The EE results showed a significant Optical flow x Session interaction ($F(2, 14) = 4.47$, $p < .05$). Post hoc Newman-Keuls comparisons revealed that in Session 2, walking at 110% PTS in Natural condition reduced energy expenditure compared to Static and N-Natural conditions, with no difference in the variability of locomotion (see Fig. 2).

In Session 3, the energetic difference between conditions disappeared, due to slight decrease in energy expenditure in Static and N-Natural conditions. This trend was accompanied by an a decrease in the variability of kinematic parameters in Static condition (stride length: $F(2, 14) = 8.83$, $p < .05$) (see Fig. 2). Finally, a significant correlation existed between energy expenditure and variability of stride length (r = 0.37), suggesting the interrelation between kinematic and energetic variables.

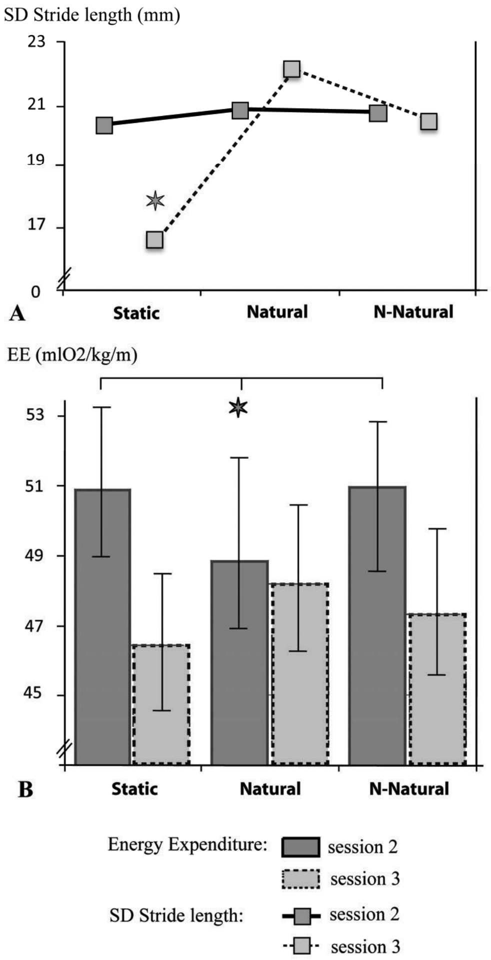


Fig. 2. *A.* Variability of stride length (SD Stride length) and *B.* Energy expenditure (EE) in each optical flow condition: Static, Natural (visual direction ≠ locomotor direction) and N-Natural (visual direction = locomotor direction) in session 2 and in session 3 (star = statistical significance, p <.05)

4.3 Discussion

Here we found that imposing an optical flow direction congruent with (i.e., opposite to) locomotion direction and three times faster than locomotor speed induced a more economic behavior. This energetic effect however disappeared with practice (session effect), suggesting that locomotor adaptations are relatively fast, characterized by a functional variability of the free parameters of walking [13].

This experiment shows that an optical flow simulating fast-speed locomotion (N-Natural condition) influences the stability and the energy consumption. Thus, these results complement previous work in the same area [9] where we found that fast optical flow increased locomotion stability and decreased energy consumption.

In conclusion for the first experiment, it appears that the *Natural* optical flow condition seems to encourage the emergence of a specific behavior characterized by a stable and economical walking pattern. In order to investigate more fully this energy-kinematic relation, a second experiment was performed with more trials per session in order to evaluate the adaptations over time.

5. Experiment 2

5.1 Procedure

Eight adults with normal or corrected vision and no motor disorder participated in the second experiment, which consisted in four experimental sessions. The first session was identical to Session 1 of the previous experiment, and was a familiarization session. The three following sessions involved one and only one of the three optical flow condition: *Static, Natural and N-Natural* conditions. The order of sessions was randomized. Each session consisted of four trials of six minutes (one visual condition). Participants were instructed to walk at fast speed (110% PTS). Trials were separated by four minutes of rest.

5.2 Results

The main result of Experiment 2 was a significant Optical flow x Trial interaction ($F(6, 42) = 3.01$, $p < .05$). Post hoc Newman-Keuls comparisons indicated that the dynamic optical flow conditions (Natural, N-Natural) produced constant energy expenditure during the four trials of each session whereas in the Static condition, energy expenditure increased between trials (see Fig. 3). Thus, walking in Natural and N-Natural conditions was more efficient than walking in the Static condition but only between the two last trials. No difference was found between the two dynamic optical flow conditions. Although (differences in) efficiency did not correlate with locomotor variability, the evolution of EE with trials was accompanied by a change in the ankle-knee relative between trials ($F(3, 21) = 6.77$, $p < .05$).

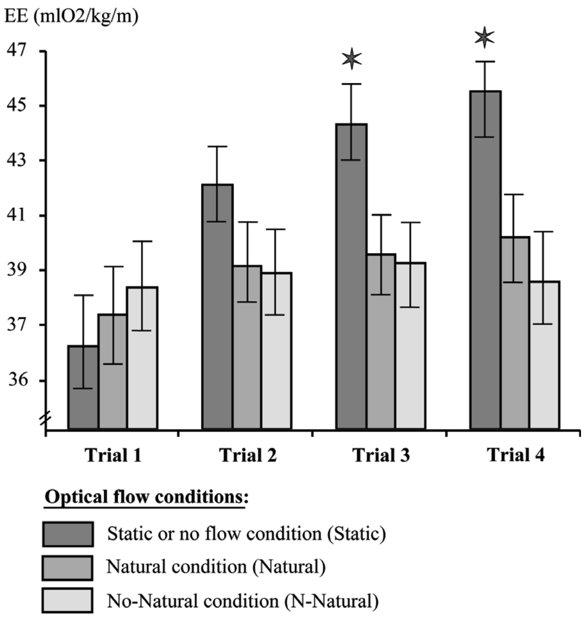


Fig. 3. Energy expenditure (EE) calculated in each trial (4 trials) of each experimental session/optical flow condition: Static, Natural and N-Natural (star = statistical significance, p <.05)

5.3 Discussion

The results of this second experiment reveal that the presence of optical flow encourages the emergence of the most economic behavior. However, contrary to Experiment 1, the energetic effect only appears from Trial 3 and is preserved until the end of the session. Moreover, this effect exists for the two optical flow conditions (*Natural and N-Natural*). The direction of the visual projection doesn't seem to play a major role, and the important parameter seems to be the presence (*Natural and N-Natural*) or absence (*Static*) of dynamic components.

6. General Discussion

These experiments reveal that interacting with optical flow during locomotion seems to produce the stabilization of the most efficient behavior, via adequate kinematic adaptations. Also, they show that optical flow modulates the energy-kinematic relation in time. Optical flow seems to be able to specify the energetic component of locomotion, or at least the kinematic-energy relation, as revealed in recent work [14].

Together, these results reveal the existence of an interrelation between many systems usually analysed separately [15]; such as the visual system, the locomotor system and the cardio-respiratory system.

Acknowledgments

This research was supported by SKILLS, an Integrated Project (IST contract #035005) of the Commission of the European Community.

References

[1] B.G. Bardy, W.H. Warren, and B.A. Kay. Motion parallax is used to control postural sway during walking. *Experimental Brain Research*, 111, pages. 271-282, 1996.

[2] W.H. Warren, B.A. Kay, W.D. Zosh, A.P. Duchon, and S. Sahuc. Optic flow is used to control human walking. *Nature Neurosciences*, 4, pages. 213-216, 2001.

[3] B. Baumberger, M. Flückiger, and R. Martin. Walking in an environment of moving ground texture. *Japanese Psychological Research*, 42, pages. 237-249, 2000.

[4] J. Konczak. Effects of optic flow on the kinematics of human gait: a comparison of young and older adults. *Journal of Motor Behavior*, 26, pages. 225-236, 1994.

[5] T. Prokop, M. Schubert, and W. Berger. Visual influence on human locomotion. *Experimental Brain Research*, 114, pages. 63-70, 1997.

[6] B.J. Mohler, W.B. Thompson, S. Creem-Regehr, H.L. Pick, and W.H. Warren. Visual flow influences gait transition speed and preferred walking speed.

Experimental Brain Research, 181, pages. 221-228, 2007.

[7] J.J. Gibson. *The Ecological Approach to Perception*, Boston: Haughton Mifflin, 1979.

[8] W.H. Warren. Action modes and laws of control for visual guidance of action. In O. G. Meijers & K. Roths (Eds.), *Movement behavior: The motor-action controversy*, Amsterdam: North-Holland, pages. 339-380, 1988.

[9] P. Guérin, and B.G. Bardy. Optical modulation of locomotion and energy expenditure at preferred transition speed. *Experimental Brain Research*, in press, 2008.

[10] F.J. Diedrich, and W.H. Warren. Why change gaits? Dynamics of the walk-run transitions. *Journal of Experimental Psychology: Human Perception and Performance*, 21, pages. 183-202, 1995.

[11] P.E. Di Prampero. The energy cost of human locomotion on land and water. *International Journal of Sports Medicine*, 7, pages. 65-72, 1986.

[12] J. Brisswalter, and D. Mottet. Energy cost and stride duration variability at preferred transition gait speed between walking and running. *Canadian Journal of Applied Physiology*, 21, pages. 471-480, 1996.

[13] K.M. Newell, and D.M. Corcos. *Variability and motor control*. Champaign, IL: Human Kinetics, 1993.

[14] D.R. Proffit. Embodied perception and the economy of action. *Perspectives on Psychological Science*, 1, pages. 110-122, 2006.

[15] R. Shaw, and J. Kinsella-Shaws. Could optical “pushes” be inertial forces? A geometro-dynamical hypothesis. *Ecological Psychology*, 19, pages. 1-15, 2007.

Investigating the role of movement in the constitution of spatial perception using the Enactive Torch

Lorenzo Grespan* Tom Froese* Ezequiel A. Di Paolo* Anil K. Seth*
Adam Spiers† William Bigge*
(*) Centre for Computational Neuroscience and Robotics, University of Sussex, UK
(†) Bristol Robotics Laboratory, University of Bristol, UK
E-mail: l.grespan@sussex.ac.uk

Abstract

This paper reports an exploratory study designed to clarify whether the Enactive Torch, a custom-built minimalist distance-to-tactile perceptual supplementation device, can be used to investigate the role of embodied action in the perception of external spatiality. By constraining the kind of exploratory movements available to the participants, we create an experimental setup in which it is possible to study the relationship between bodily degrees of freedom and spatial perception. We present a preliminary investigation of the strategies used by minimally trained participants to locate various objects placed in front of them by engaging in active exploration under constrained conditions.

1. Introduction

Since the early 1990s there has been a growing consensus within the cognitive science community that the body shapes the mind [7, 4]. At present, the challenge is to build on this general consensus by further explicating the specific contribution of embodiment to our mental capacities. One particular focus of interest in this respect is the constitutive role of embodied action for perception [10, 11, 15]. The study of this kind of 'enactive' perception is greatly facilitated by the use of novel technological interfaces [9, 2], especially sensory substitution devices (also known as "perceptual supplementation" devices for reasons given in [8]).

Already in 1969 Bach-y-Rita and his colleagues employed a vision-to-tactile system called TVSS, as "a practical aid for the blind and as a means of studying the processing of afferent information in the central nervous system" [17]. They demonstrated that active exploration with a TVSS, which essentially consists of a camera hooked up to an array of tactile stimulators located somewhere on the body, allowed trained blind subjects to perceive the world as if seen through a camera. Moreover, some subjects spontaneously reported the experience of an externalisation of the stimulation on their body into the world that is in many respects similar to vision.

This seminal study opened a vivid debate about the phenomenology of perception enabled by the use of perceptual supplementation devices, which still continues in the cognitive sciences today [6, 13, 10, 3, 16]. However, despite four decades of research into perceptual supplementation devices, as well as a growing fascination with the phenomenological aspects of their usage, so far no consensus has been reached on how to best understand this type of technology. Indeed, there are ongoing disagreements about some of the most fundamental issues, especially in terms of whether the afforded perception is (i) essentially an extension of the substituting perceptual modality, (ii) the constitution of percepts in the substituted modality, or even (iii) the constitution of a new way of perceiving that is dependent on the specific kind of sensorimotor profile provided by the technological interface [1]. This situation is made even worse due to the fact that the proponents of competing theories often cannot even agree on the experiential phenomenon, i.e. what it is like to use a perceptual supplementation device, that is to be explained.

Some preliminary steps toward the development of a pragmatic phenomenological research program that could address these difficulties were reported by Froese and Spiers [6]. They introduced the *Enactive Torch* (ET), a minimalist perceptual supplementation device, precisely for this purpose. Here we complement those efforts by testing whether this device is also a suitable tool for psychological experiments, especially for investigating the role of embodied action in the constitution of spatial perception. In this paper, thus, we shall revisit some some work originally done by Lenay and Steiner [9], by using the ET. The main objective, apart from testing the original results using a different setup, is to put the ET to a more rigorous experimental test in order to identify the advantages and potential problems with this new scientific tool.

2. The Enactive Torch

In response to the lack of agreement about fundamental issues pertaining to perceptual supplementation technology, Froese and Spiers [6] developed the Enactive Torch (ET), a minimalist device that has been designed to be cheap, non-intrusive as well as easy to use. Accordingly, the ET has the potential of becoming a widely distributed research tool within the scientific community, and thereby help to move the seemingly open-ended debate about the nature of perceptual supplementation forward. In particular, its aim is to inform the specification of the phenomenology of using perceptual supplementation devices by more easily giving researchers first-person access to the experiences in question, an essential source of insight that has so far been sorely lacking in this debate, as well as in the cognitive sciences more generally [14]. A second-generation prototype of the ET is shown in Figure 1 below.

Figure 1: A second-generation prototype of the Enactive Torch (courtesy of A. Spiers).

The main body of the ET contains the power source (batteries) and the circuitry; the separate handle is equipped with an ultrasonic sensor mounted on its end, a small servo-motor with a rotating disc and a vibrotactile actuator. The vibro-tactile motor can generate a set of vibration patterns of variable intensity that can be felt by gripping the handle. In its normal mode of operation the strength of vibration/angular displacement of the disc is proportional to the distance of the closest object in the ultrasonic sensor's range.

In this work, the servo-motor is inactive; we only made use of the vibro-tactile output. The ET is employed in 'binary mode', i.e. the strength of the response can only assume all-or-nothing values according to whether or not an object is present in the ET's field sensor range. The maximum range in this mode of operation is limited to approximately 60 cm; objects are detected if localised within a cone of aperture ca. 30°.

3. The experiment

This study is inspired by the work of Lenay and Steiner [9] who used a minimalist interface to investigate aspects of perceptual awareness. Their interface was composed of a single photo-electric cell that triggers a binary tactile stimulator whenever the incident luminosity within a cone of about 20° is greater than a specific threshold value. Even with such a simple device the localisation of luminous targets is still possible through active exploration. Moreover, it was found that the perception of depth requires a greater capacity for action than the detection of a target object's orientation in relation to the body of the participant. The authors thus argue that the space of lived experience is co-extensive with the space of action and perception, and that the perception of objects does not occur separately 'behind' the perceiver's point of view, but rather in the very same space in which the perceiver moves.

We implemented a set of experiments specifically designed to replicate the work by Lenay and Steiner with a novel perceptual supplementation device: the ET was combined with a simple controlled environment which allows participants to explore the experimental setup with 1 and 2 degrees of freedom (DoFs). The purpose of this study is to measure to what extent blindfolded participants can perceive the position of a target in a novel environment. It is organised into two tasks:

Task 1: Participants are asked to detect the horizontal displacement of a target object by moving the ET horizontally along a fixed 1D axis (one DoF).

Task 2: Participants are asked to detect the distance to a target object by a combination of horizontal movement and rotation of the ET about its centre (two DoFs).

The additional degree of freedom in the second task provides a basis for the participants to perceive distance in addition to an object's horizontal displacement.

3.1. Methods and materials

The experimental setup consists of a sliding platform placed on top of a 160 cm long rail (see Figure 2). Objects are placed in the test-space in front of the rail. The ET is mounted onto the platform, thus being constrained to horizontal movement. The participants can move the platform to both ends of the rail by extending their arm and, if needed, sliding along with their chair. The possibility of using the rotational DoF by turning the platform is only enabled for the second task.

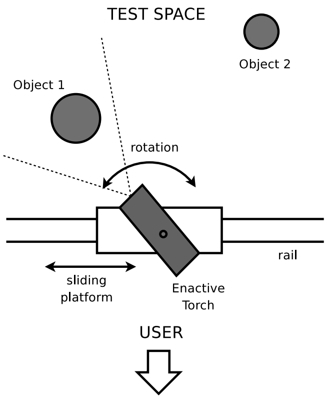


Figure 2: Experimental setup. The user (indicated by the large arrow) can perceive the presence of an object within the ET's range by sliding and rotating the platform on top of which the device is mounted.

The objects are classified according to their size as small (3 cm), medium-sized (9 cm) and large (32 cm), and the distances with respect to the rail are classified as near (8 cm), medium (16 cm) and far (42 cm). We make use of one small and two medium-sized cylindrical objects, as well as one large flat object (a 'wall').

The conic shape of the sensor's receptor field gives rise to an 'inverse shadow' effect: the farther the object, the larger it appears (see Figure 3). This effect could potentially be used by participants as a criterion to distinguish between near and far objects, solely on the basis of their apparent horizontal length.

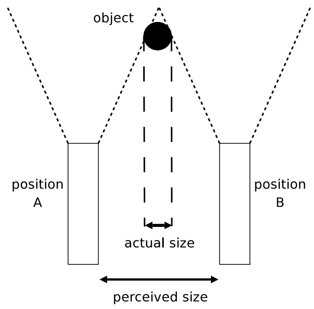


Figure 3: The 'inverse shadow' effect. Assuming the device is moving from left to right, the user will start perceiving the object at position A and stop perceiving it at position B. The object's apparent size is bigger than its actual size.

To prevent participants from perceiving an object's distance as a simple function of its apparent size, we provided them with an experimental situation that cancels out the regularity of the 'inverse shadow' effect through ambiguity. Accordingly, the size of the 'wall' has been chosen such that, when placed at a near distance, it appears as large as a small object placed on the far end. Moreover, the 'wall' is flat to discourage participants from paying attention to its depth, encouraging them instead to distinguish the object only by means of its length (during the second task).

Since the ET was set to operate in binary mode and all the objects were placed within the range of its receptor field, the tactile response triggered by the presence of an object does not indicate that object's distance. Hence, we expect that the minimum number of DoF for which perception of distance can emerge is two.

3.2. Participants

Sixteen participants volunteered to take part in this study, mostly researchers in the fields of Informatics and Psychology. Participant mean age is 31.66 with standard deviation 12.77; there are two outliers (age 65 and 45). In the pool of participants 43.75% (7 participants) are women, 87.5% (14 participants) are right-handed and 31.25% (5 participants) are non-native speakers of English. Two of the participants were already familiar with the ET device; all participants received preliminary training until a basic level of competence was achieved. The experimental tasks were novel to all of the participants.

3.3. Experimental protocol

The number of trials per training task (marked by * in the list below) ranges from three to five, depending on a participant's ability. The participants are blindfolded while carrying out these trials, but they are allowed to visually verify their answers after each trial (this was not allowed during Tasks 1c and 2b to avoid implicit training). At least one trial of Task 2a involved the 'wall' as well as a small object placed at different distances. In this manner we provided participants with the experience that they could not rely on the 'inverse shadow' effect in order to make judgements about distance. The training process allowed them to acquire a sufficiently good mastery of the experimental setup on average within three minutes of first using the ET.

1. Task 1: sliding only (1 DoF)
   (a) * count number of objects (2-4)
   (b) * determine the wider of 2 objects
   (c) determine the centre of 1 object
2. Task 2: sliding and rotation (2 DoFs)
   (a) * determine the further of 2 objects
   (b) evaluate distance to 1 object

For Task 1c participants were asked to explore the target space by moving the ET horizontally, and to place the ET pointing in the (perceived) centre of the object. At this point the experimenter takes a measurement of the correctness, replaces the object, and asks the participant to perform another trial. During Task 2b the participant was asked to declare the object's distance (either 'near' or 'far'). The experimenter records the answer, replaces the object, and asks the participant to perform another trial. The participants remained blindfolded during both tasks, and no form of verbal feedback was given by the experimenter. More details about the objects used in the trials is given in Table 1.

Task	# objs.	obj. size	obj. distance	# trials
1a	2-4	any	any	3-5
1b	2	different	same	3-5
1c	1	any	any	3
2a	2	different	different	3-5
2b	1	any	any	4

Table 1: Experimental parameters. An object size of 'any' indicates that they could have a different or same size; an object distance of 'any' means they were placed at a random distance of 'near', 'medium' or 'far'.

While carrying out the experiments, we took measures to minimize contextual clues - such as the noise produced by placing objects in the test-space - that could potentially be exploited by participants to answer successfully by means other than those intended.

4. Results

Task 1c was achieved successfully by all participants in every trial. Reported centres generally differed by no more than 1 cm from actual centres. Since the ultrasonic sensor is slightly inconsistent across object shapes and textures, an average error of 1 cm is reasonable and we did not feel it necessary to look for greater accuracy.

In terms of Task 2b, the experimental results show that two DoFs are sufficient for the participants to detect the distance of target objects (see Table 2). Participants correctly classified the object distance as 'near' or 'far' in 81.25% of the cases (standard deviation is 0.2627). Those participants who reported using a specific strategy to solve the task correctly classified distance in 88.1% of the cases (standard deviation is 0.1597).

Most participants reported attempts to generate a strategy to carry out the task at hand, in particular with respect to the more elaborate Tasks 2a and 2b. We distinguished these reported strategies into three categories: (i) *cognitive*, (ii) *intuitive*, or (iii) *unknown*. In

participants	accuracy	std. dev
all	81.25%	0.26%
with strategy	88.10%	0.16%

Table 2: Results of Task 2b. The figures in the first row include two participants who reported not having developed any way to solve the task.

category (i) we placed all approaches that are based on some explicit geometric/analytical thinking. In these cases, once a strategy has been developed, the participant generally tries to carry it out as if performing the steps specified by an algorithmic procedure. In category (ii) we placed those approaches that rely on some kind of intuitive feelings or pre-reflective bodily skills. Here, the participants judged the success of their embodied actions in terms of a felt sensation. The last category (iii) includes those participants that reported not being aware of any way of solving the task, and who thus resorted to guessing. Table 3 shows the results for each of these categories in terms of Task 2b.

Category	# participants	accuracy	std. dev
cognitive	11 (68.75%)	90.91%	0.15
intuitive	3 (18.75%)	77.78%	0.16
unknown	2 (12.5%)	33.34%	0.33

Table 3: Results of Task 2b for different categories of behavioural strategies. The cognitive strategies are significantly better than the intuitive strategies.

Most participants (68.75%) reported to have used a cognitive strategy, and this strategy turned out be significantly better when compared to the intuitive strategies. Note that the two participants in the 'unknown' category performed worse than chance level (50%), though this is likely due to the small number of trials. We will now describe the behaviour involved in the cognitive and intuitive strategies in more detail.

Cognitive strategies. During Task 1c, when the ET was limited to only sliding along the rail, there was only one type of strategy that was reported. First, the participant would explore the space, until the object was detected. Second, the width of the object is scanned by slowly moving the platform at a constant speed. This provides the participant with a rough estimate of the length of stimulation experienced while the device 'traverses' the whole object. Finally, the participant backs up for half the length of stimulation and then stops. This should leave the ET pointing near the centre of the object. Although only four participants explicitly reported having adopted this method, it has been observed in others as well.

With respect to body movement/posture, it is interesting to note that all participants tried to minimize the number of moving parts: three subjects were observed to only move the arm and keep the rest of the body still, whereas three others kept the arm fixed in position and moved the body instead (by sliding their chair horizontally). One participant reported using the elbow as a marker for the location of an object's earliest detection since the first training session (Task 1a); the marker was later exploited to yield an estimate of the object's size, as per Task 1b.

The extra DoF in Tasks 2a and 2b enabled a broader range of strategies in comparison to the approaches developed for the 1 DoF tasks. The most frequent strategy (observed in seven participants) consisted in (1) pointing the ET at the centre of the target object, and (2) rotating it in both directions until the object was out of the sensor's range. While behaviour (1) essentially consisted in the strategy reported for Task 1c, behaviour (2) made it possible to detect the distance of an object because nearby objects would generate longer stimulation during rotation than far objects.

Another approach relied on the 'inverse shadow' effect to give an estimate of distance as a function of perceived size. Even though all subjects were aware that this regularity holds only for same-shape objects, four participants reported using it as an initial estimate, switching then to the rotational strategy as a method for validation. In particular, by rotating the ET participants were able to detect if the target was the 'wall' - and if not so, then the inverse shadow effect can be thought to hold to some extent. When the target object was indeed the wall, the task took significantly longer to complete, albeit consistently with the correct answer (i.e. it is a 'near' object).

Another interesting example comes from a participant whose approach in Task 2a consisted in (a) positioning the ET about halfway between the two objects, then (b) rotating the handle until the first target object was out of range, and finally (c) repeating (b) for the second object. The extent of rotation needed for stimulation to cease was used as an inverse correlate of object distance, and the objects could be related to each other as 'near' or 'far'. In Task 2b this participant used an approach which consisted in sliding the ET horizontally in one direction until it reached the end of the object; at this point the ET would be slid and rotated in the opposite direction, tracing out tangents to the object. The rate of decrease of the angle was used as a source of information to determine the object's distance.

Finally, one participant devised a strategy for Task 2b based on positioning the ET near one of the ends of the sliding rail. The ET is then rotated, up to a 90° angle, until the device detects the object. If empty space is detected between the object and the rail, the object is reported as being 'far'.

Intuitive strategies. Three participants reported using some kind of "feeling" in order to achieve the tasks, and were unable to provide a step-by-step description of their behavioural strategy. They reported finding it difficult to move the device at a constant speed during Tasks 1b and 1c, as well as having to rely on "sensations" to estimate the width of objects. Interestingly, one of them reported visualising an "imaginary space" for this task.

Regarding Tasks 2a and 2b, the three participants stated that they were unable to find an explicit strategy for distance estimation, and thus decided to rely on the intuitive sense of distance that was generated through their exploratory actions rather than on some form of analytical thinking.

Note that significantly less participants adopted an 'intuitive' strategy. Considering that under non-experimental circumstances, such as when using the ET to find your way across a room, most people report having an intuitive sense of their spatial environment within minutes of their first exploratory activities, this is slightly odd. We speculate that this discrepancy could be the result of a more general problem, namely the attempt to study enactive perception under controlled and minimalist conditions, even though such unnaturally constrained situations are more likely to elicit detached problem-solving attitudes in the participants. If this is indeed the case, then this field would be confronted with even more significant methodological problems than previously assumed.

5. Discussion

It is worth noting that some of the participants spontaneously made observations about their experience that could be interpreted as indicating the beginnings of the constitution of spatial perception. For example, during Task 2b one participant reported that using the ET for active exploration with two DoFs was like "projecting my consciousness forward", while another said that it "felt like being able to see around the object". It appears that, at least for these two participants, their activity of using the ET might have indeed resulted in them being intentionally directed into the world through an actively constituted perception of space.

Interestingly, and contrary to our expectations, these two participants also reported using a behavioural strategy that is based on cognitive inference rather than on skilful tool manipulation. However, it is possible that they conceptualised their pre-reflective behaviour in this manner after being prompted to explain their actions by the experimenter (cf. [5]). In future work it will be essential to record the actual movements of the participants in order to be able to classify their strategies according to objective movement, and not just verbal reports. Such a study would also benefit from better means of collecting phenomenological reports, perhaps by means of second-person interview methods (e.g. [12]). Accordingly, the analysis of strategies presented here should only be viewed as preliminary.

Future work could use the ET to investigate Dreyfus and Dreyfus's [5] description of a progression from novice to expert in a task. While we typically follow rules when we are getting used to a novel task, we become less reliant on such rules as we become more skilful. If behavioural strategies could be objectively classified into cognitive and intuitive, then this suggests a future study where the developmental progression of one class of strategies into the other could be investigated.

Acknowledgements

The authors are grateful to the participants who volunteered to take part in the experiment. They would also like to extend their thanks to all those who contributed to building the experimental scenario and refining its concept, as well as to Nele Froese for providing extensive feedback on a draft of this paper.

References

[1] M. Auvray and E. Myin. Perception with compensatory devices: From sensory substitution to sensorimotor extension. *Cognitive Science*, in press.

[2] M. Auvray, D. Philipona, J. K. O'Regan, and C. Spence. The perception of space and form recognition in a simulated environment: The case of minimalist sensory-substitution devices. *Perception*, 36(12):1736–1751, 2007.

[3] N. Block. Tactile sensation via spatial perception. *Trends in Cognitive Sciences*, 7(7):285–286, 2003.

[4] A. Clark. *Being there: Putting brain, body, and world together again*. MIT Press, Cambridge, MA, 1997.

[5] H. L. Dreyfus and S. E. Dreyfus. Peripheral vision: Expertise in real world contexts. *Organization Studies*, 26(5):779–792, 2005.

[6] T. Froese and A. Spiers. Toward a phenomenological pragmatics of enactive perception. In *Proc. of the 4th Int. Conf. on Enactive Interfaces*, pages 105–108, Grenoble, France, 2007. Association ACROE.

[7] S. Gallagher. *How the body shapes the mind*. Oxford University Press, New York, NY, 2005.

[8] C. Lenay, O. Gapenne, S. Hanneton, C. Marque, and C. Genouëlle. Sensory substitution: Limits and perspectives. In Y. Hatwell, A. Streri, and E. Gentaz, editors, *Touching for Knowing: Cognitive psychology of haptic manual perception*, pages 275–292. John Benjamins, Amsterdam, The Netherlands, 2003.

[9] C. Lenay and P. Steiner. Externalism and enaction. In *Proc. of the 4th Int. Conf. on Enactive Interfaces*, pages 145–148, Grenoble, France, 2007. Association ACROE.

[10] A. Noë. *Action in Perception*. MIT Press, Cambridge, MA, 2004.

[11] J. K. O'Regan and A. Noë. A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5):939–1031, 2001.

[12] C. Petitmengin. Describing one's subjective experience in the second person: An interview method for the science of consciousness. *Phenomenology and the Cognitive Sciences*, 5(3-4):229–269, 2006.

[13] J. Prinz. Putting the brakes on enactive perception. *Psyche*, 12(1):1–19, 2006.

[14] F. J. Varela and J. Shear. First-person methodologies: What, why, how? *Journal of Consciousness Studies*, 6(2-3):1–14, 1999.

[15] F. J. Varela, E. Thompson, and E. Rosch. *The embodied mind: Cognitive science and human experience*. MIT Press, Cambridge, MA, 1991.

[16] P. B. y Rita. Sensory substitution and qualia. In A. Noë and E. Thompson, editors, *Vision and Mind: Selected Readings in the Philosophy of Perception*, pages 497–514. MIT Press, Cambridge, MA, 2002.

[17] P. B. y Rita, C. C. Collins, F. A. Saunders, B. White, and L. Scadden. Vision substitution by tactile image projection. *Nature*, 221(5184):963–964, 1969.

The Role of Visual Feedback when Grasping and Transferring Objects in a Virtual Environment

Andrea H. Mason* Brandon J. Bernardin*
(*)University of Wisconsin-Madison, USA
E-mail: amason@education.wisc.edu, bernardin@wisc.edu

Abstract

This work explored how people use visual feedback when performing simple reach-to-grasp and object transfer tasks in a tabletop virtual environment. Visual feedback about the index finger and thumb was provided in three conditions: vision available throughout the movement, vision available until movement onset, or vision absent throughout the movement. We hypothesized that vision available until movement onset would be an advantage over no vision for all tasks. However, results indicated that the role of vision was influenced by task such that visual feedback up to movement onset was only beneficial in the reach to grasp task. These results suggest that the role of visual feedback for performance in virtual environments is task specific.

1. Introduction

Along with the technological advances seen over the past 20 years in the field of Virtual Reality (VR), challenges to the widespread implementation of virtual environments (VEs) have surfaced. In particular, realistically implementing graphical effects in interactive VEs has been problematic, leading to distortions in perception (i.e. swimming of the visual world, lack of motion parallax, misperception of depth) and degraded performance (i.e. disturbed perceptual-motor coordination) when users try to perform both simple and complex tasks [1-6]. This is particularly true for the presentation of accurate feedback about hand movements during interaction in VEs. In order to address this problem, a systematic study of the use of sensory feedback in VEs is required.

In natural environments, it has been shown that when humans perform simple aiming or reach-to-grasp movements with full vision of their limbs they perform these tasks efficiently and accurately [see 7 for a review]. With no visual feedback, movements tend to become slower and less accurate than when vision is present [8-11].

In contrast to the rich visual world available in natural environments, in VEs feedback must be synthetically created in order to be presented to users. As such, it is appealing to determine whether performance can be optimized with decreased fidelity of sensory information or by reducing the amount of time that the feedback is made available. To date, studies on the role of visual feedback in virtual environments have indicated that a representation of one's self is advantageous for performance [12-18]. However, it has also been shown that the timing of the presentation of that feedback can be manipulated while still leading to effective performance. In a recent series of studies, Mason & Bernardin [14,15] manipulated the timing of visual feedback in a simple reach, grasp and lift task by removing feedback of the hand for the entire trial, presenting feedback prior to movement onset, presenting feedback up to peak movement velocity or presenting it after peak velocity. In this series of studies, the hand was represented as two small spheres located at the tips of the index finger and thumb. Results indicated that visual feedback provided before movement onset and up to peak reaching velocity resulted in movement times that were similar to when the representation of the hand was available throughout the entire trial. In contrast, feedback that was presented after peak velocity resulted in performance that was similar to a no feedback condition. Mason and Bernardin [14,15] argued that vision during the first third of the movement improves performance because it can serve to calibrate proprioceptive knowledge of one's position. Specifically, when visual feedback is provided before movement onset or early in the movement, the sensory feedback from the visual system is compared and calibrated with sensory information from the proprioceptive system. Thus, the more accurately calibrated proprioceptive system can be used to successfully complete the movement after vision has been removed.

It is important, however, to keep in mind that the role of environmental variables during performance is task specific [see 16 for a review]. As such, the same sensory information can be used differently when we perform different tasks. In the current study object transfer from hand to hand and person to person was investigated to elucidate whether the above results regarding the role of visual feedback about one's own movements extend to other tasks beside reaching to grasp a stationary object. Thus, we tested whether the advantage of having vision available prior to movement initiation extends to tasks such as object transfer between hands and between people. We hypothesized that, like simple reach to grasp movements, having a representation of oneself before movement initiation would facilitate object transfer when passing objects between hands and between people.

2. Method

2.1 Subjects

Thirteen right-handed subjects participated in a single 30-minute experimental session. Two subjects were members of the research lab. The mean age of the participants was 26±5 years. Subjects provided informed consent before participating and ethical approval was obtained from the University of Wisconsin-Madison Social and Behavioral Institutional Review Board.

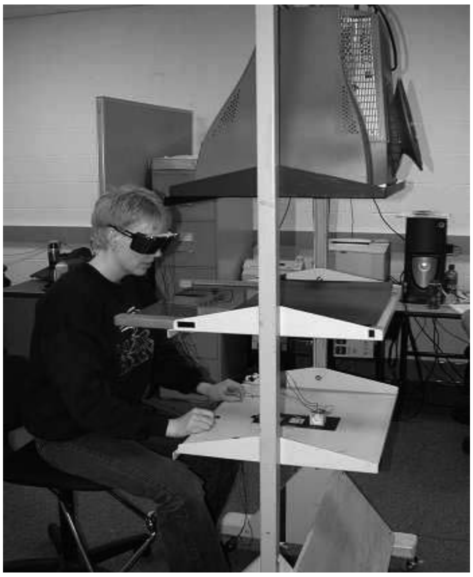


Figure 1. Experimental apparatus

2.2 Experimental apparatus

Data was collected in the Wisconsin Collaborative Virtual Environment. Shown in Figure 1, a graphic image of a target cube was displayed on a downward facing monitor. A half-silvered mirror was placed parallel to the computer screen, midway between the screen and the table surface. The image on the computer screen was reflected in the mirror and appeared to the subject to be located in a workspace on the table surface.

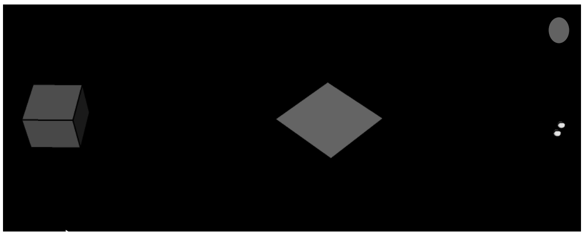


Figure 2: Subject's view of the virtual workspace. The target cube is shown on the left, the red diamond represents the interception zone, the small yellow spheres represent the index finger and thumb.

Stereoscopic images were achieved by alternately displaying the left and right eye images on the monitor at 60Hz and synchronizing these images with the CrystalEYES™ goggles worn by the subject. Three light-emitting diodes (LEDs) were fixed to the goggles and were tracked by a VisualEyez 3000 motion capture system (Phoenix Technolgies, Inc.) at a sampling rate of 200 Hz. This information was processed with a 10 ms lag, to produce a head-coupled, stereoscopic view of the virtual environment on the work surface.

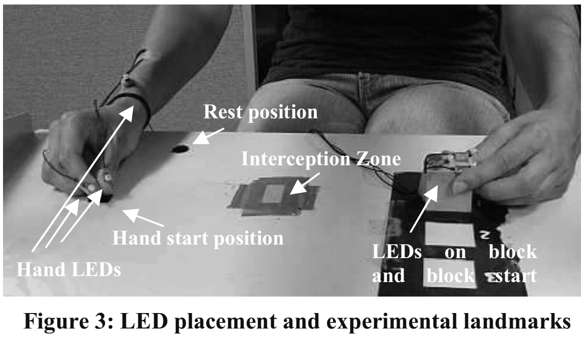


Figure 3: LED placement and experimental landmarks

LEDs were positioned on the subject's left and right index fingers, thumbs and wrists (see Figure 3). In some conditions, the tracked positions of the LEDs on the right hand were used to generate a crude graphical representation of the hand in the form of yellow spheres (3 mm diameter) superimposed on the tips of the index finger and thumb (see Figure 2). Three LEDs were positioned on the top surface of the wooden target cube, which measured 3.2 X 3.2 X 3.2

cm (see Figure 3). The tracked positions of LEDs on the physical target cube were used to generate a superimposed graphical object, which had identical dimensions to its physical counterpart.

2.3 Experimental design and Procedure

The graphical view seen by the subject is shown in Figure 2. Physical landmarks in the real workspace are shown in Figure 3. Subjects were prevented from seeing the real workspace via the placement of a shield below the mirror. Thus subjects did not have visual feedback of their moving limb or the surrounding environment.

Subjects began each trial with the pads of the right index finger and thumb lightly pinched together on a haptic “rest” position (Velcro™ tab on the table surface). The appearance of a green light in the upper right hand corner of the work surface was the subject’s cue to move from the “rest” position to the start position (another Velcro™ tab located approximately 3 inches from the “rest” position). No graphic representation of either mark was available; subjects found these marks based on touch alone. Once the fingers remained stationary on the start position for one second, the light turned from green to red in conjunction with the appearance of the target cube. The goal for the subject was to reach medially 15 cm with the right hand to grasp the target within a graphically presented interception zone (3.5 cm X 3.5 cm) under three task conditions and three visual feedback conditions. In the first task (Reach to Grasp: RG), the target was placed on the work surface within the interception zone. The subject simply reached for, grasped, lifted and moved the stationary target back toward the right start position. In the second task (Transfer) the subject held the object with the left hand at a start position located 15 cm to the left of the interception zone. The subject’s task was to transport the object from the left start position while also reaching for the object with the right hand such that the object could be transferred from the left to the right hand within the interception zone. In the final task (Pass), the object was held by the experimenter at the start position on the left side. The experimenter transported the object while the subject reached from the right start position so that the object could be passed from the experimenter to the subject within the interception zone. Analysis of the block velocity profiles indicated that the experimenter performed the passing movements consistently over all trials in the experiment. For each of these tasks, subjects were presented with three visual feedback conditions. In the first feedback condition, subjects did not receive a representation of their right hand throughout the trial (no vision; NV). In the second feedback condition, a graphical representation of the right hand was available prior to movement initiation, but was extinguished as soon as the subjects lifted their hand off of the start position (vision extinguished at movement onset; VEX). In the third feedback condition, subjects received a graphical representation of their right hand throughout the entire movement (full vision; FV). In all conditions, graphical information about the size and location of the target was available throughout the trial, however graphical feedback about the position of the left hand on the target cube was not provided.

These manipulations resulted in a balanced design of 3 *tasks x 3 visual feedback conditions.* Subjects performed ten blocked trials in each condition for a total of 90 trials. Conditions were presented in a random order for each subject to minimize the effects of rank.

2.4 Data Analysis

Kinematic measures of human performance were employed to evaluate the temporal and spatial aspects of the movements. Specifically, movement time was used as a general index of task performance, peak reaching velocity was used as a measure of open-loop processes and motor planning and percent time from peak reaching velocity was used as a measure of closed-loop control. To quantify grasp formation we used peak hand aperture (resultant distance between the index finger and thumb) and the timing of peak aperture.

To extract the dependent measures described above, the position data recorded by the VisualEyez motion analysis system was smoothed with a 7 Hz low-pass second-order bi-directional Butterworth filter. A customized computer program (KinSys, Eh?Soft) was used to determine the start of movement based on a criterion wrist velocity of 5mm/s. The end of movement was determined visually as the point immediately preceding the second rise in wrist velocity. This second rise is associated with the movement of the block toward the start position once a stable grasp has been achieved. End of movement minus start of movement was used to calculate movement time (MT). The position data were differentiated and peak resultant velocity (PV) and the timing of that peak (TPV) were extracted. Percent time from peak velocity was defined as (MT-TPV)/MT*100. The resultant distance between the index finger and thumb was calculated and both Peak Aperture (PA) and the timing of peak aperture (TPA) were extracted. Percent time from peak aperture was defined as (MT-TPA)/MT*100. Data were analyzed using separate repeated measures ANOVAs and an a priori alpha level of $p < 0.05$. Means and standard error measures are reported for significant results.

3. Results

3.1 Reach Kinematics

As illustrated in Figure 4, the interaction between task and visual feedback condition was significant ($F_{4,48}$=2.5, p=0.05) for movement time. The effect of visual feedback on movement time was further explored by conducting separate ANOVAs for each task. Visual feedback had a significant influence on movement time for the transfer ($F_{2,24}$=4.7, p<0.05) and reach to grasp ($F_{2,24}$=11.6, p<0.001) conditions, however the effect was not significant for the pass condition. Adjusted pairwise comparisons in the reach to grasp task indicated that movement time in the NV condition was significantly longer than movement times in both the FV (p<0.01) and VEX (p<0.01) conditions. The FV and VEX conditions were not significantly different from each other. Pairwise comparisons in the transfer task indicated that movement time in the FV condition was significantly shorter than in the NV (p<0.01) and the VEX (p<0.05) conditions. The NV and VEX conditions were not significantly different from each other. It is also important to note that the interaction between task and visual condition just failed to reach the level of significance (p=0.076) for %TFPV, however the trends for this measure mirror the results seen for movement time. No main effects of condition were found for the magnitude of peak velocity or percent time from peak velocity.

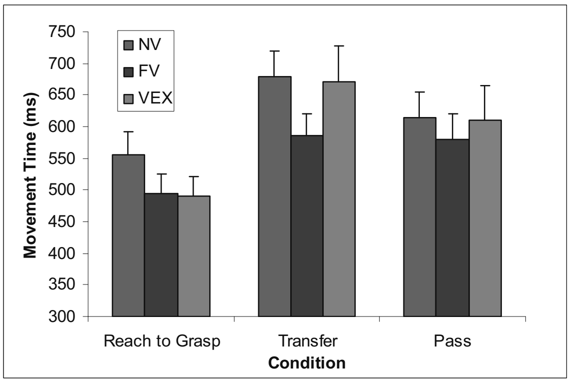


Figure 4: Means and standard errors for the interaction between task and visual feedback condition for MT.

As shown in Figure 5, a main effect of visual feedback condition was found for percent time from peak velocity ($F_{2,24}$=5.4, p<0.05). Adjusted pairwise comparisons indicated that deceleration time was longer in the NV condition when compared to the FV condition (p<0.01) only.

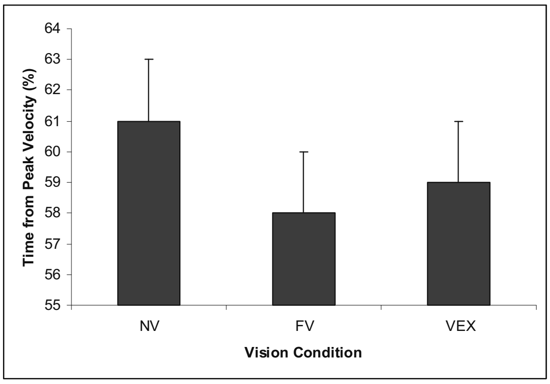


Figure 5: Means and standard errors for the main effect of task on percent time from peak velocity

3.2 Grasp kinematics

The interaction between task and visual feedback condition was significant ($F_{4,48}$=5.4, p<0.01) for peak aperture. Figure 6 illustrates this interaction. Simple main effects indicated that visual feedback had a significant influence on peak aperture for the reach to grasp ($F_{2,24}$=21.0, p<0.001), transfer ($F_{2,24}$=5.4, p<0.01) and pass ($F_{2,24}$=13.9, p<0.001) conditions. Pairwise comparisons in the reach to grasp task indicated that all means were significantly different (p<0.05). Apertures were smallest when full vision was present, larger when vision was provided up to movement initiation and largest in the NV condition. The results of pairwise comparisons in the transfer task indicated a significant difference between apertures in the Full Vision (FV) and No Vision (NV) conditions (p<0.05) only. Finally, the comparisons in the pass task indicated that apertures were larger in the NV condition when compared to the FV and VEX conditions (p<0.01).

A main effect of task was found for percent time from peak aperture ($F_{4,48}$=11.00, p<0.01). Pairwise comparisons indicated that percent hand closure time was longer (p<0.01) for the reach to grasp task (41 ± 2 %) than the pass task (36 ± 2). None of the other comparisons were significantly different. Finally, a main effect of vision was found for percent time from peak aperture ($F_{2,24}$=15.5.00, p<0.01) which indicated that percent hand closure time was shorter in the FV (35.7 ± 2%) condition when compared to both the NV (41.7 ± 2%, p<0.001) and VEX (38.3 ± 2%, p<0.05) conditions.

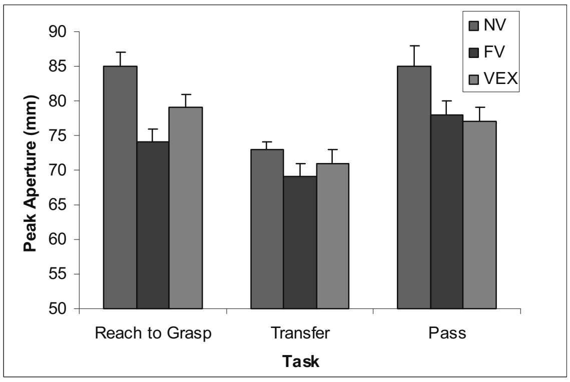


Figure 6: Peak Aperture means and standard errors for the interaction between task and visual feedback condition.

4. Discussion

Our purpose was to investigate the role of visual feedback about one's own movements when reaching to grasp a target, transferring the target between one's own hands or grasping the target from another person.

While a representation of one's own movements is important for optimal performance in goal directed aiming and reach to grasp tasks [12-15,17-18], it has been noted that visual feedback that is extinguished at movement onset can also be used to calibrate the proprioceptive system leading to effective target acquisition [9, 10, 14, 15]. Specifically, Mason & Bernardin [14] found that movement times in a reach to grasp task were similar for a full vision condition when compared to a condition in which feedback was extinguished at movement onset. Results of the current experiment replicated [14] for the reach to grasp task, however these results did not extend to the transfer and pass tasks. While subjects benefited from the limited visual feedback provided in the VEX condition when performing the reach to grasp movement, that feedback was not sufficient in the object transfer condition. For the transfer task, movement times were as slow in the VEX condition as the NV condition. Why the difference between the reach to grasp and transfer tasks? We hypothesize that the complexity of the transfer task necessitated the use of online visual feedback available at the end of the movement for homing in on the target. Specifically, in a simple reach to grasp task visual feedback available only at the beginning of the movement may suffice because a complete plan to grasp the stationary target can be generated before the movement is initiated. This plan will then be executed using vision calibrated proprioceptive feedback. However in a transfer task, where the object to be grasped is moving, online updates to the plan are necessary as the position of the target changes during performance. Because visual feedback was not available in the VEX condition, subjects were forced to rely on less accurate proprioceptive feedback for executing these modifications to the movement plan.

For the pass task a different pattern of results emerged. Here, movement times were similar regardless of visual feedback condition. Why did the presence or absence of visual feedback not influence performance in this task as it did in the reach to grasp and transfer tasks? It has been shown that when receiving an object from a partner, subjects are sensitive to the motion of the object they are grasping [18, 19]. Specifically, they use characteristics of the object's motion to generate their own movement. Thus, subjects may have attended to visual feedback about the moving target instead of visual feedback about their own hand.

Results for grasp scaling indicated a different role for visual feedback about the hand for this component of the movement. Larger grasp apertures have typically been associated with more complex tasks that demand greater attentional resources [20]. It is believed that a larger aperture is used as a compensatory strategy to avoid missing or hitting the target. Similar to results found by Mason & Bernadin [15] when subjects simply reached to grasp a target, grasp apertures were smallest in the full vision condition, slightly larger for the VEX condition and largest when vision was not available. These results indicate that although the visual information available in the VEX condition allowed for some improvement in aperture formation, compensations for the lack of feedback after movement onset were made to ensure an accurate grasp. Thus, movement complexity was significantly reduced when visual feedback was made available throughout the task. Similar results were seen in the transfer task where apertures were smaller in the full vision condition when compared to the no vision condition. In contrast, for the pass task, apertures for the FV and VEX conditions were similar and smaller than apertures in the NV condition. When considered along with the movement time results, this result suggests that subjects may have used visual information about the position of the fingers prior to movement initiation to generate a plan for more accurate grasp scaling despite the fact that this visual information was not used during the execution of the movement.

In conclusion, feedback is an essential component of any environment and providing too much (i.e. distracting), too little or incorrect feedback can have significant deliterious effects. Thus, determining how and when to provide that feedback is important for creating a successful VE. The results presented here suggest guidelines for the provision of sensory information about hand motion during interaction tasks:

1. A representation of hand movement is not always required throughout the entire task. Specifically, reach to grasp tasks with low precision requirements can benefit from a representation provided only before movement onset. Grasping an object from another person may not require a representation of the moving hand.
2. The requirement for sensory information is not only task specific but also specific to components within a task and as such it is important to understand the nature of the task when deciding on the presentation of sensory information.

5. References

[1] R.S. Allison, L.R. Harris,M. Jenkin, M., U. Jasiobedzka, J.E. Zacher. Tolerance of temporal delay in virtual environments. In Proc. VR 2001, IEEE Computer Society Press (2001), 247.

[2] F. Biocca, J. Rolland. Virtual eyes can rearrange your body: adaptation to virtual eye location in see-thru head-mounted displays. Presence-Teleoperators and. Virtual Environments 7, (1998), 262-277.

[3] P. DiZio, J.R. Lackner, Circumventing side effects of immersive virtual environments. In M.J. Smith, G. Salvendy, R.J. Koubek (Eds). Advances in Human Factors/Ergonomics. Vol 21. Design of Computing Systems, Amsterdam: Elsevier (1997), pp. 893-897.

[4] J.Y. Jung, B.D. Adelstein, S.R. Ellis. Predictive compensator optimization for head tracking lag in virtual environments. In Proc IEA 2000/HFES 2000, (2000) pp. 499-502.

[5] C. Ware, R. Arsenault. Frames of reference in virtual object rotation. In APGV '04: Proceedings of the 1st Symposium on applied perception in graphics and visualization, ACM Press, New York (2004), pp. 135–141.

[6] C. Ware, R. Balakrishnan. Reaching for objects in VR displays: lag and frame rate. ACM Transactions on Computer.-Human Interaction 1 (1994), 331–356.

[7] L.G. Carlton. Visual processing time and the control of movement, in: L. Proteau, D. Elliott (Eds.), Vision and Motor Control, North-Holland, New York, 1992, pp. 3-31.

[8] R.S. Woodworth. The accuracy of voluntary movement. The Psychological Review: Monograph Supplements 3 (1899) 1-114.

[9] N.E. Berthier, R.K. Clifton, V. Gullapalli, D.D. McCall, D.J. Robin. Visual information and object size in the control of reaching. Journal of Motor Behavior. 28 (1996) 187-197.

[10] M. Gentilucci, I.Toni, S. Chieffi, G. Pavesi. The role of proprioception in the control of prehension movements: a kinematic study in a peripherally deafferented patient and in normal subjects. Experimental Brain Research. 83 (1994) 447-482.

[11] L.S. Jakobson, M.A. Goodale. Factors affecting higher-order movement planning: A kinematic analysis of human prehension. Experimental Brain Research. (1991), 199-208.

[12] V. Buchman, T. Nilson, M. Billinghurst. Interaction with partially transparent hands and objects. *In Proc. AUIC 2005*, IEEE Computer Society Press (2005), 17-20.

[13] E.D. Graham, C.L. MacKenzie. Physical versus virtual pointing. In Proc CHI 1996, ACM Press (1996), 292-299.

[14] A.H. Mason, A.H., B.J. Bernardin. The role of early or late graphical feedback about oneself for interaction in virtual environments. Proceedings of the Laval Virtual Reality International Conference. (2007) 121-130.

[15] A.H. Mason, A.H., B.J. Bernardin. Vision for performance in virtual environments: The role of feedback timing. Journal of Sport and Exercise Psychology 30 suppl (2008) S63.

[16] R.A. Magill Motor control and learning: Seventh Edition. McGraw Hill, New York, New York, 2004.

[17] A.H. Mason, M.A. Walji, E.J. Lee, C.L. MacKenzie. Reaching movements to virtual reality and graphic objects in virtual environments. In Proc CHI 2001, ACM Press (2001) pp. 426-433.

[18] A.H. Mason, C.L. MacKenzie. The effects of visual information about self-movement on grip forces when receiving objects in an augmented environment. In Proc IEEE VR, Haptics Symposium, IEEE Computer Society Press (2002) 105-112.

[19] A.H. Mason, C.L. MacKenzie. The role of graphical feedback about self-movement when receiving objects in an augmented environment. Presence, Teleoperators and Virtual Environments 13 (2004) 507-519.

[20] A.M. Wing, A. Turton, C. Fraser. Grasp size and accuracy of approach in reaching. *Journal of Motor Behavior 18* (1986) 245-260.

6. Acknowledgments

This work was supported by the National Science Foundation under Grant No. 0346871. We would also like to thank Sarah Ferrito and Scott Mason for data collection and software support.

The Learning Effect of Force Feedback Enabled Robotic Rehabilitation of the Upper Limbs in Persons with MS - a Pilot Study.

De Boeck J.* Alders G.† Gijbels D.† De Weyer T.*
Raymaekers C.* Coninx K.* Feys P.†
(*) Hasselt University - tUL - IBBT, Expertise Centre for Digital Media, Wetenschapspark 2, B-3590 Diepenbeek, Belgium
(†) REVAL Rehabilitation and Health Care Research Centre, PHL University College, Guffenslaan 39, B-3500 Hasselt, Belgium.
E-mail: {joan.deboeck, tom.deweyer, chris.raymaekers, karin.coninx}@uhasselt.be
{galders, dgijbels, pfeys}@mail.phl.be

Abstract

Multiple Sclerosis (MS) is an autoimmune disorder resulting in different sorts of physical dysfunctioning, such as loss of limb function, in-coordination, altered muscle-tone, etc. Within limits, these symptoms can be treated by rehabilitational measures such as strength and functional training. The intensity of the rehabilitation is one of the key factors to a possible success. To increase the intensity, rehabilitation robotics can be a promising new technology. In this study, a Phantom haptic device was applied during a force feedback enabled training program focussing on the upper extremities in persons with MS. Seen the fact that we found no significant learning effect during the first contact with the environment and seen the improvements of the upper limb performance after 4 weeks of robotic training, this pilot study shows that force feedback supported rehabilitation can be a promising emerging new therapy. However, further research is needed to refine the technology behind it and to explore the full potential of the patient's enactive knowledge while transferring training effects of the computer generated environment to daily life functional capacity.

Keywords: Robotic Rehabilitation, Multiple Sclerosis, upper extremity, Task Performance

1 Introduction and Related Work

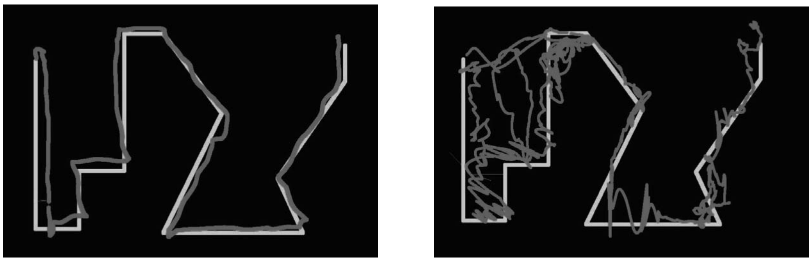
Multiple Sclerosis (MS) is an incurable chronic and progressive disorder of the central nervous system, resulting in secondary symptoms such as impairments of strength, muscle tone, sensation, co-ordination, balance, as well as visual and cognitive deficits. These symptoms, caused by an autoimmune response, progressively lead to severe limitations of functioning in daily life, while still no final cure exists. Besides, the recent use of medication only focussed on slowing down or reducing the worsening of the disease or the symptoms. Therefore, still an important part of the therapy consists of physical and occupational training and exercises

Evidence can by found in literature that the intensity and duration of a rehabilitation session are key factors for its efficiency [6]. Indeed, studies of exercise therapy focusing on balance and walking parameters, have shown a beneficial effect with regard to e.g. muscle strength [15]. From this point of view, over the last years, several research projects have been conducted in order to provide the patients with (virtual) environments that can be used in a more independent way at a level customised to his or her abilities. Ultimately, this approach should open perspectives for the patients to practice at home under the (remote) supervision of a therapist [9, 7].

Some of these solutions, such as the Rutgers Ankle rehabilitation interface [2] or the MIT-MANUS project [5] , apply force feedback technology successfully into rehabilitation training. It must be stated that these research projects were mostly conducted on hemiparethic stroke patients. Although dysfunctions caused by stroke, which is not a progressive disease, improve naturally as time progresses from the cerebro-vascular

(a) The car-trajectory task (b) The object manipulation task (c) The speeded-tapping task

Figure 1: Screenshots of the three applications



(a) Resulting path of a healthy person (b) Resulting path of an MS patient

Figure 2: Resulting paths of some example trials of the car-trajectory game

incident on, these studies reported promising results in speeding the rehabilitation progress and augmenting outcome-level in these patients compared to controls. Therefore the use of haptic rehabilitation technology could be useful in a degenerative disease as MS, as well. The application of force feedback can be considered to be part of one of the following classes [8]:

- Passively moving the patient's limbs
- Actively assisting the patient's movements by restricting false movements and facilitating correct motions.
- Actively resisting the patient's movements, evoking higher forces from the patient for the execution of the movement.

This creates an opportunity for the therapist to set up a personal training plan, according to the patient's physical abilities and/or needs. Moreover, the available haptic feedback provides the patient with a very direct (first order) feedback loop, also stimulating the patient's sense of proprioception [1].

2 Objectives

Very few studies have properly investigated the therapeutic potential of the integration of force feedback in upper limb training in persons with MS. In this project a virtual environment (VE), in the form of three simple games, has been realised, stimulating MS patients to improve their skills. Both visual and haptic feedback is presented to the user while performing these tasks. The performance of execution and its results are permanently logged.

The overall aim of the investigation was to assess the potential of a force feedback enabled VE as a training tool in the context of the rehabilitation of lost upper limb function of persons with MS. We took into account any improvements concerning the specific tasks in the VE, but also possible functional improvements in everyday tasks, as well as a potential increase of the overall muscle strength of the patient's arm. Moreover, the motivational aspect (how well are patients motivated to practice with the proposed setup during an entire training period) was another point of uncertainty during the training program, lasting four weeks.

3 Training Environment

The training environment consisted of a standard desktop PC, with speakers and a 19" CRT monitor. The Phantom 1.5 haptic device was used by the patients to control the training tasks. To reach the abovementioned objectives, a virtual environment was created, consisting of three training applications. Taking into account that some MS patients have reduced cognitive abilities and to minimise initial learning effects, the tasks have to be simple and easy to learn. We have chosen for a *trajectory* task, an *object manipulation* task and a *speeded-tapping* task (see figure 1).

In the first task, the patient had to steer a virtual car over a pre set trajectory. The patient was aided in this task by restricting the movements of the Phantom to a 2D plane in which the road was located. An adjustable force was applied attracting the car to the centre of the road. This spring force could be set in 3 levels, ranging from small to medium and large, each changing the spring stiffness with a factor 10. The actual spring constants were empirically determined during several pretests.

For the second task, users had to pick up a (virtual) book lying on a shelf and they had to place it in a bookcase. The applied forces simulated the gravity and inertia of the book. According to the patient's capabilities, the weight of the book (force feedback setting) could be adjusted in 3 levels from 0.5, 1.0 and 1.5 kilograms.

The third task was a virtual implementation of the well-known speed tapping task (Eurofit Test Battery) [14]. For this task, a guiding force which restricted the patient's movement to a vertical plane could be set. As in the other two tasks, an incremental 3-level force feedback adjustment available: the first level created two stiff walls where the cursor was kept in between. The second level implemented a spring force centering the cursor, and the third level provided no guiding plane at all.

Finally, according to the patients' needs, the scaling factor between the real and the virtual movements (Control Display Gain) could be adjusted. A large scale required larger but less accurate movements, while a small scale appeared to be more difficult due to the required accuracy.

4 Experimental Setup

The experimental setup was approved by the local ethics committee of the University of Hasselt and by the ethics committee of the Rehabilitation & MS-centre Overpelt. The inclusion criterion for the MS-group was a dysfunction of the arm due to muscle weakness. The exclusion criteria were a relapse of MS or treatment with corticosteroids in the last month prior to the study, upper limb paralysis, severe cognitive dysfunction and severe visual dysfunction. After their written agreement 21 persons with MS (from the Rehabilitation & MS-centre Overpelt) were included in this study (13 female and 8 male, mean age 59,7 $\pm$ 1,16 years). Additionally, ten healthy subjects, selected among the rehabilitation centre's personnel (n=10, 4 female and 6 male, mean age 48,00 $\pm$ 6,5 years) participated in a 'healthy control' group. The experiment lasted for a duration of nine weeks in total. In the first week, all patients participated in the intake sessions. Subjects had to pass several functional tests (Nine Hole Peg Test, Purdue Pegboard test, ARAt and TEMPA [13, 4, 12, 3]); also their maximum force at the upper extremities was measured (JAMAR handgrip force and MicroFet isometric muscle testing (Biometrics, Gwent, UK)) and an EMG/Accelerometer measurement was done while performing simple everyday tasks (combing hair, pouring water in a glass, reaching for an object). For the healthy control group normative data, available in literature, was used for all of these tests, except for the EMG-accelerometer test.

After the completion of these tests, all subjects (MS and healthy controls) were asked to complete the three virtual applications in random order. For each of the three tests four subsequent trials were completed in the same day. The first trial was used to familiarise the subjects with the haptic device. The second trial was the first trial that was logged. For further analyses this trial was named *'Initial 1'*. The subsequent trials were named *'Initial 2'* and *'Pre'* respectively. This last trial was used to represent the baseline measurement. The first two measurements were added to evaluate the occurring learning effect.

The force feedback and scale settings for all these trials were standardised to the same values for all subjects. After the intake, the MS-group participated in individual training sessions during a period of four weeks. The training volume was three sessions per week during 30 minutes per session in which the intensity of the training was augmented per week at the same level for all subjects. This training frequency is in accordance with the ACSM Guidelines for Exercise Testing and Prescription for elderly people[11]. During each half hour training, patients were guided by an occupational therapist to rehearse every task as much as possible. At the first training session of each following week, the first trial performed for each application was logged using the standardised settings. Afterwards, the force feedback and scale settings were adjusted for al training trials in that week to the same level for al participants. During the training sessions, the MS-group was exposed to different variations for the trajectory task and the object manipulation task. These variations were a change of the form of the track or a change of endposition of the book.

After the training period of four weeks, all patients had to complete the functional tests, the force measuring and the EMG/accelerometer tests again, after which one trial of each of the applications was logged for the last time (post measurement). Four weeks after the last exposure to the experiment, 11 persons with MS were subjected to a follow-up test for the three applications.

For all the experiments, the total execution time and total travelled distances were logged to an SQL database at 200Hz. With the tapping task, additionally the number of correct and false taps was logged. During the execution, other parameters such as the current position, actual velocity, force and deviation (distance) from the ideal path were also logged.

5. Results

In the analysis presented in this paper, only the first 3 measurements (Initial 1, Initial 2, Pre) as well as one measurement after a training period of 4 weeks (post) are taken into account. The difference between a healthy person and an MS patient is immediately visible from the 'Pre'-test data. Figure 2 illustrates the position profile of the car trajectory task. The straight light line represents the ideal path where the car is supposed to stay on the road. The darker curved line shows the actual path driven.

Figure 3 gives a view on the velocity profiles of a tapping task for a typical healthy person and an MS patient. For the healthy person, we see a velocity profile as expected according to the findings of [10]. The velocity profile can be approximated by a (skewed) parabola. With the MS patients, we see a similar profile, but as if there is an additional sine superimposed either in the accelerating side, the decelerating part, or in both.

Figure 4 shows the *average completion time* for the trajectory task and the object manipulation task and the *average velocity* of the tapping application for both the MS group and the healthy control group for the three subsequent logs *initial 1*, *initial 2*, *pre*. This data allows us to draw conclusions on the initial learning effect after a maiden exposure to the virtual tests/tasks and before starting the training program.

Analysing the different parameters using ANOVA for repeated measurements (completion time, total distance, average and peak velocity), globally we find only no significant to small learning effects, except for the following numbers:

- The completion times for the MS group in the trajectory task points out a significant reduction between initial 1 and initial 2: p=0,023485
- For the tapping task, there is a significant raise of velocity for the MS group between initial 1 and initial 2: p=0,028
- The velocities of all the trials of the Tapping task for the control group raised significantly from initial 1 to the pre measurement: p=0,003.

Therefore, it seems that 3 trials are enough to get a stable performance on the tasks, developed in this study. This way, the insignificant learning effects were only of short duration and a basic level of performance was reached from whereon the implemented training period could start.

Almost all parameters of the control group were significantly better than those of the MS group, which can be considered as a 'double check' test for the relevance of the selected patients.

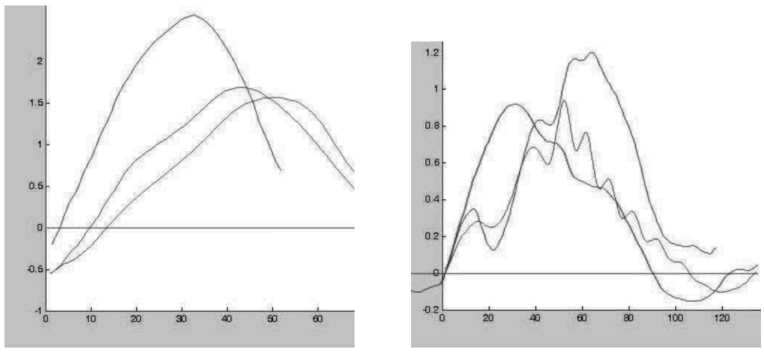
Comparing the same parameters after the training period of four weeks the results for the MS group are given in figure 5. For the Trajectory Task there is a significant improvement between the first trial and the post measurement (p=0,00025), between the second trial and the post measurement (p=0,014) and a borderline trend between the pre en the post test (p=0,062). For the Object Manipulation task we measure a significant improvement over all other measurements (p<0,0004). For the tapping task however, we could not measure any significant training effect.

6 Discussion

As we measured for the MS group only very small (but not significant) learning effects which are in line with the learning effects of the healthy control group, we may conclude that the overall difficulty of our experimental task is suitable for the targeted MS group. We hence have a very 'easy to learn training task', which was one of our objectives. At least this is true for the first two applications. Consequently, we can conclude that the usage of the Phantom as a haptic training device causes no significant problems for the patients to adapt to. Nevertheless, it has to be noted that we observed some severely affected patients having more difficulties manipulating the Phantom in front of their body during the entire duration of a training session, which often led to extremely compensating poses and muscle fatigue.

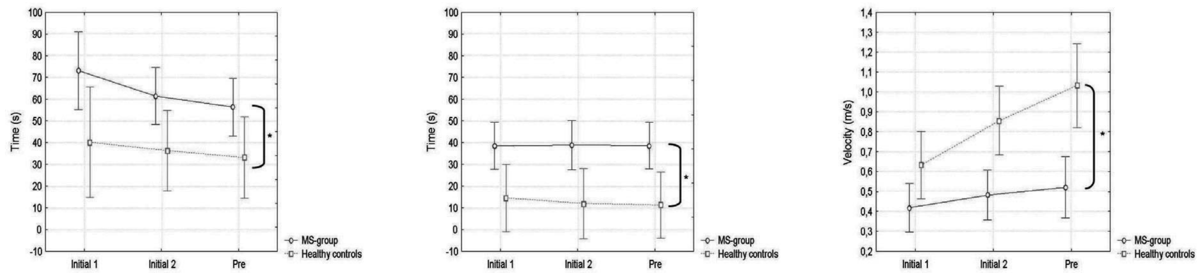
On the other hand, for the trajectory task and the object manipulation task, we were able to measure a significant training effect after 4 weeks. As there was no significant learning effect, we may conclude that the improvement was the result of the training sessions and not merely an effect of better understanding of the task, which was one of the intended goals.

The results are less straightforward for the tapping task. For the MS group, we see a slight (but insignificant) trend towards higher velocities (figure 4(c)). The post test, however, shows no significant improvement at all. On the contrary, some subjects even performed worse! Analysis of other recorded parameters, training data and a comparison with the results of the functional tests will be necessary to explain this outcome. Besides,



(a) Result of 3 typical healthy subjects (b) Result of 3 typical MS patients

Figure 3: Velocity profile of the tapping task of a typical tap



(a) Initial Learning Effect: Average **Completion Time** for the Trajectory Task (b) Initial Learning Effect: Average **Completion Time** for the Object Manipulation Task (c) Initial Learning Effect: Average **Velocity** for the Tapping task

Figure 4: Results of the first three trials of each application. ($*p \leq 0.05$)

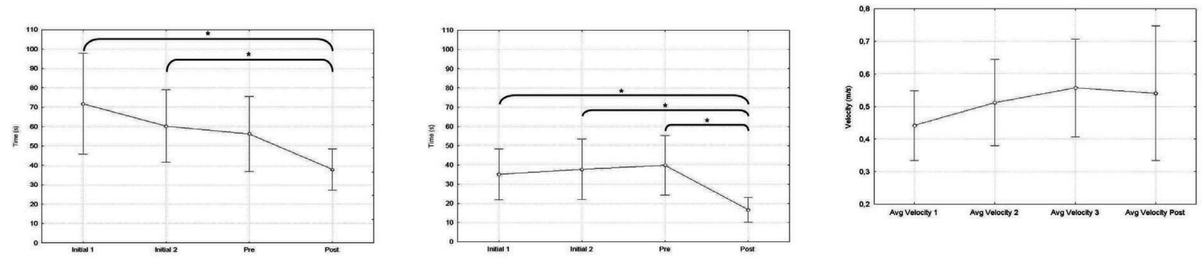
it is remarkable that unlike the other experiments, the subjects in the healthy control group also need a learning period for this task. This makes us conclude that the virtual tapping task behaves somewhat different than the other two tasks.

Another important aspect, which was one of the indications to apply a force feedback enabled training approach, is the patient's motivation. Indeed the relation between the virtual reality task (robot training) and the 'gaming' aspect and the feeling during an according real life task, were enormous motivational aspects for the patients to participate in this experiment. However, due to the limited variations in the different tasks, after a few weeks, patients reported verbally to get tired of always having to complete only slightly varying tasks in the VE. For optimal exploitation of the patient's motivation, more alternatives, or a progressive 'gaming scenario' seems to be necessary.

Finally, in a preliminary analysis, we found no correlation between the patients' performance in the virtual tasks and their muscle strength. This is somewhat surprising, as the Phantom is manipulated in front of the body using all frontal shoulder muscles. Patients reported that the upper limb muscles got tired during the robotic training. Therefore we may suppose that a four-week training session creates a light training stimulus for the patient's arm muscles but that it could not be measured by means of the analysis made at this moment. Although the training interval (3 times a week during half an hour) but not the duration (4 weeks) is conform with the recommendation of the ACSM guidelines for elderly people, further analysis of the data is necessary to confirm this assumption.

7 Conclusion and Future Work

From this pilot study, we may conclude that the training protocol supported by force feedback, beholds beneficial effects concerning the level of performance of the upper limbs for the particular virtual task. According to the found learning effects (Initial 1, Initial 2, Pre), the use of this system should make it possible for the patients to start an exercise programme quite quickly. The level of difficulty at the start of the training period appears to be suitable for this specific group of intended users and the applied force feedback creates the opportunity to start exercising gradually, adjusting the intensity/difficulty according to the capabilities of the patient.



(a) Training Effect After 4 weeks: Average **Completion Time** For the Trajectory Task

(b) Training Effect After 4 weeks: Average **Completion Time** for the Object Manipulation Task

(c) Training Effect After 4 weeks: Average **Velocity** for the Tapping task

Figure 5: Result of the training effect in a) trajectory task, b) object manipulation task, c) tapping task in the MS group after four weeks (last column), results are mean values ± SD ($*p \leq 0.05$)

Future research will be needed to further explore the full potential of robotic upper limb rehabilitation. Also new and different sorts, and a more extensive amount of virtual tasks need to be developed in order to keep this kind of rehabilitation appealing to its users and to further explore the possible transfer to activities of daily living.

8 Acknowledgements

Part of the research at EDM is funded by the ERDF (European Regional Development Fund) and the Flemish government. The REVAL and the EDM contribution in this research were funded by the INTERREG-III programme (project Nr. 4-BMG-II-1=84, 01/09/2007 - 31/05/2008, Euregio Benelux) with the added TRANSCEND funding (01/02/2008 - 31/05/2008, Euregio Meuse-Rhine), both Interregional European funding organisations. The authors of this paper would also like to thank all patients and healthy subjects that participated in this study, as well as the therapists and staff of the Rehabilitation & MS-Centre in Overpelt, Belgium. Special thanks go to Mia Thijs and Niels Goyens from this centre.

References

[1] C. Cadoz. Concepts for enactive interfaces. Presentation at the first Enactive Workshop 21–22 march, 2005.

[2] J. E. Deutsch, J. Latonio, G. C. Burdea, and R. Boian. Post-stroke rehabilitation with the rutgers ankle system: A case study. *Presence: Teleoper. Virtual Environ.*, 10(4):416–430, 2001.

[3] P. Feys, M. Duportail, D. Kos, P. Van Asch, and K. P. Validity of the tempa for the measurement of upper limb function in multiple sclerosis. *Clin Rehabil*, 16(2):166–173, 2002.

[4] J. Gallus and V. Mathiowetz. Test-retest reliability of the purdue pegboard for persons with multiple sclerosis. *Am J Occup Ther*, 57(1):108–111, 2003.

[5] H. Krebs, M. Ferraro, S. Buerger, M. Newbery, et.al. Rehabilitation robotics: pilot trial of a spatial extension for mit-manus. In *Journal of NeuroEngineering and Rehabilitation*, October 26 2004.

[6] G. Kwakkel, R. van Peppen, R. Wagenaar, S. Dauphinee, C. Richards, et. al. Effects of augmented exercise therapy time after stroke: a meta-analysis. In *Stroke 35*, pages 2529–2539, 2004.

[7] G. Lathan. Dimensions of diversity in design of telerehabilitation systems for universal usability. In *CUU '00: Proceedings on the 2000 conference on Universal Usability*, pages 61–62, New York, NY, USA, 2000. ACM.

[8] P. Lum, D. Reinkensmeyer, R. Mahoney, W. Rymer, and C. Burgar. Robotic devices for movement therapy after stroke: Current status and challenges to clinical acceptance. *Top Stroke*, 8:40 – 53, 2002.

[9] R. M. Telerehabilitation. In *NeuroRehabilitation 12*, pages 11–26, 1999.

[10] D. Meyer, R. Abrams, S. Kornblum, C. Wright, and J. Smith. Optimality in human motor performance: Ideal control of rapid aiming movements. pages 340–370, 1988.

[11] A. C. of Sports Medicine. *ACSM's Guidelines for Exercise Testing and Prescription*, volume 7th edition. Lippincott Williams and Wilkins, Philadelphia, 2006.

[12] T. Platz, C. Pinkowski, F. van Wijck, I. Kim, P. di Bella, and J. G. Reliability and validity of arm function assessment with standardized guidelines for the fugl-meyer test, action research arm test and box and block test: a multicentre study. *Clin Rehabil*, 19(4):404–411, 2005.

[13] E. Rosti-Otajarvi, P. Hamalainen, K. Koivisto, and L. Hokkanen. The reliability of the msfc and its components. *Acta Neurol Scand*, 117(6):421–427, 2008.

[14] L. Vanhees, J. Lefevre, R. Philippaerts, M. Martens, W. Huygens, and T. T. How to assess physical activity? *Eur J Cardiovasc Prev Rehabil*, 12(2):102–114, 2005.

[15] D. R. White LJ. Exercise and multiple sclerosis. In *Sports Med 34*, pages 1077–1100, 2004.

Cognitive Cooperation in Virtual Reality: an Haptical Enabled Experiment

Paolo Tripicchio* Emanuele Ruffaldi* Carlo Alberto Avizzano* and Massimo Bergamasco*
(*) Scuola Superiore S.Anna , via Martiti 11, Pisa, Italy
E-mail: p.tripicchio@sssup.it, e.ruffaldi@sssup.it, carlo@sssup.it

Abstract

Cooperation is a fundamental element of human behavior that arises in many contexts from physical activities to purely cognitive activities. In most of the cases cooperation emerges from the adoption of a strategy that can be adopted before the task execution or identified during the task itself. At the same time the cooperative process requires a communication channel for the exchange of information. In most of the cases the visual and the audio channels are sufficient for performing this cooperative action, but there are cases in which other channels are involved. This paper presents an experiment in the area of cognitive cooperation during the execution of a purely cognitive task in which the channel used for cooperation is haptic based.

1 Introduction

When coping with large-scale problems people usually tend to resolve them as a team where members cooperate to achieve a common goal. Usually in such process team members are also decentralized to make use of specific resources. The cooperation can happen in a coupled way, like in a production line or in a loosely coupled way like in a robot football team[8]. What is important for the team's work effectiveness is the knowledge management. In knowledge-intensive teamwork, cognitive abilities and cooperation between users determine the efficiency and quality of the final performance. In general it's possible to distinguish three levels of cooperation: the work cooperation level, where team members follows their specific workflow, the information sharing level, where team members communicate to share information and the cognitive cooperation level, where team members learn from each other and use past experience and skills to solve new problems.

In literature many studies where performed to study problems of multi-agent cooperation for distributed artificial intelligence[1, 2, 3, 5, 6, 9]. Multi agent systems are systems where two or more agents interact to satisfy a goal. An agent is a locus of problem-solving activity and operates asynchronously with respect to other agents. Each agent has it's own autonomy that refers to the ability of making its own decisions about the activities to perform, when to do them and what type of information to communicate with the others. Autonomy can be limited by policies built into the agent or cause of the environment of operation. In multiagent interaction agents solve subproblems that are interdependent because of the sharing of resources or relationship among the subproblems. This is due to the natural decomposition of domain problem solving into subproblems. In particular is possible to have subproblems that are the same or overlapping (the case in our study) or have two subproblems that are part of a larger problem in which a solution exist under the assumption of certain constrains among the two subproblems. The cooperation and coordination can be further complicated when the information used to make decisions is incomplete, out of date or inconsistent. Other investigations regard cooperation, dynamic decision and creative user interfaces[7, 12].

Cognitive flow occurs during the passing of cognitive informations generated during the cooperation process between members. So as we'll depict later we will use this flow to get results on how cognitive information can enhance the performance during our experiment. Cooperation at human level is not well studied in literature compared to other species like insects, so group problem solving is a science not already fully uncovered. For example evolutionary psychology study cognition but only at individual level. Cognitive cooperation operates unconsciously without the need of learning. Cooperation is retained useful for tasks that exceed the ability of a single to be executed. In Wilson et al. [11] the authors summarize that performance is not based on learning and that the advantages of groups require being in a group.

In this paper is presented a cooperative experiment to be performed in a Virtual Reality environment. In particular the experiment make use of haptics and is intended to investigate cognitive cooperation under some assumptions. The objectives of the experiments are:

- to study and understand the interaction between perception and problem solving with visual and haptic feedback
- how cognition improves group performances

How cognition interacts when humans attempt complex tasks is not yet defined by theories. In particular problems crop up on information exchange between perception and problem solving. Perception needs to deliver information to the problem solving process and the problem solving process has to communicate directions to the perceptual process.

Next section will introduce the intended VR experiment defining the idea, it's mathematical definition and the architecture involved during the experimental sessions. Than We'll present the performing sessions and the results.

2 The Experiment

With this proposed experiment we wanted to investigate cooperation, interaction and cognition of users in a Virtual Reality shared environment. In particular under the condition of reducing the knowledge of the environment at disposal of each user.

The experiment is based around a maze escape task. The two users have to cooperate to get out the maze, scoring the highest number of points . Points are increased by collection of bonuses, decreased by contacting maluses and the total task execution time decreases the number of points. The bonuses and maluses are clearly visible and do not move in the maze.

The two users have to cooperate to perform the task because they are connected by an invisible virtual rope: the movement in the maze is the combination of the relative movements of the users. Each user control the movement through the use of a haptic interface that generates a force feedback proportional to the movement applied. If the corresponding movement is not in accordance with the desired one, it could both signify that the two users are not coordinated or that an obstacle constrains the movement itself. The decomposition of the task in the two subtasks of escaping and collecting bonuses has been divided among the two users. The first user sees only the maze while the other sees only the bonuses and the maluses as shown in Figure 2. The visual channel is used to understand the direction of movement, goals and constraints, while the intention of the user is transmitted through the haptic feedback. In this case the audio channel is not used and the users are being asked to not talk.

The difference in perception of the scenario and visual feedback for the two users can be mathematically represented in the following way. We have a set A representing the VR environment with subsets A_1 and A_2 (figure 1). Calling respectively the the first user U_1 and the second U_2 we want to investigate the interaction of

$$(U_1|A_1)|(U_2|A_2)$$

where | means conditioned.

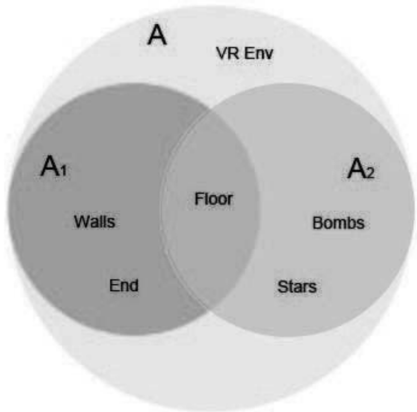


Figure 1: Chart of environment subsets

3 Architecture

The experiment has been carried out in a shared virtual environment developed within the Virtual laboratory(VL) Framework [10]. In particular the VL allows to develop rapidly shared experiments in a virtual 3d space with the benefit of haptic interaction. With the use of such technology we resolve the problem of synchronicity between user's interactions into the scene, we manage the haptic feedback directly without the need of external ad-hoc software and most of all, it's possible to retrieve, in real time, all the relevant data that we need to process to get the scientific results that we wanted to achieve. From the point of view of user interaction and force feedback we decide to use the GRAB [4] haptic interface to explore the virtual maze.The GRAB is a robotic device with a large workspace of of 400 mm depth, 400 mm high and 600 mm wide. This haptic device is able to generate continuous forces at the end-effector of 4 N in the worst condition while the peak forces are up to 20 N, the device has a position accuracy at zero load less than 1 percent,

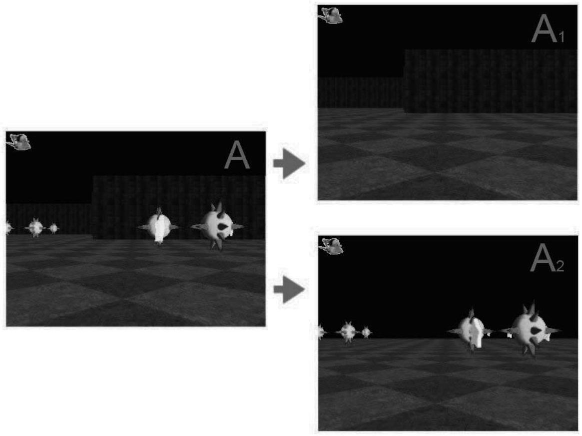


Figure 2: The maze environment as seen by the two different participants

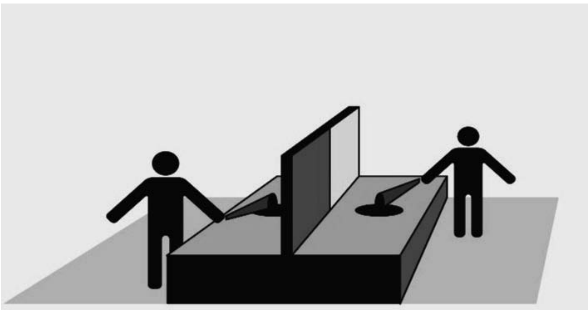


Figure 3: Setup for the experiment

i.e. 1mm over 100 mm. The force feedback was calculated as a spring plus dumper attached to the camera so that to move the camera in one direction or to rotate its visive cone the user should apply a force proportional to the desired movement. We removed the fingertip at the end effector cause our manipulation metaphor required an interaction with the whole hand. In particular the two users participating to the experiment can't see each other and can't talk to reach the goal (figure 3) so that the performance was strictly related to cognitive informations.

4 Experiment Setup

We organized an experiment for evaluating the cooperative concept discussed above. In particular 10 volunteers, from 22 to 31 years, 8 male and 2 females, mostly students with previous computer experience, have participated to the experiment. The users were asked to reach the goal position in the maze with the highest number of points, where the number of points was affected mostly by the completion time, than by the number of bonuses collected and the maluses avoided. The shape of the maze was the same for all the experimental sessions and it is shown in figure 6 in which the red dots are the bombs (maluses), and the yellow ones are the stars (bonuses).

The participants have been divided in 5 groups of two and none of them have experienced this system before. Each group performed one session with 4 trials each, performing the same task with the following variations. The first trial(T_1) is the reference trial in which every user is able to see only one part of the scene. That corresponds to

$$(U_1|A_1)|(U_2|A_2)$$

. The second trial(T_2) is aimed at understanding the labyrinth and for that reason the two users see all the environment

$$(U_1|A)|(U_2|A)$$

. The third and fourth trials are used for assessment, in particular the third trial has the same conditions of the first, in the fourth the users' roles are swapped:

$$(U_1|A2)|(U_2|A_1)$$

. The aim of the fourth trial is to check if visual memory influences the performance or if it is the previously obtained knowledge to guide the interaction.

The aim of these trials was to obtain performance data related to knowledge and cognition. In particular, to check the improvements gained from cognitive flow, we asked two of the groups that had performed the experiment to explain the task and possible strategies to the following group. In particular the groups 1, 3 and 5 have performed the task without any instruction while the groups 2 and 4 have received the explanation from the previous group.

After the experimental session we asked the users to fill a questionnaire that has been designed to identify the impressions of the users respect the creation of strategies. In particular the questionnaire asked them the easiness of performing the task and the impression of the time needed for completing it. In addition for every trial the presence of a strategy have been asked, and when applicable if the suggestions of the previous group were useful for the creation of a strategy.

Next section will depict the results obtained by such experiments. During the tests many questions could be aroused in the cognitive process, in particular every time the 3D view wasn't moving even if the user had applied a force to move it, users should try to understand why it happened: was the other user applying a force in the opposite direction and with the same force or there was an obstacle on the maze not visible that couldn't permit to move in the desired direction? This

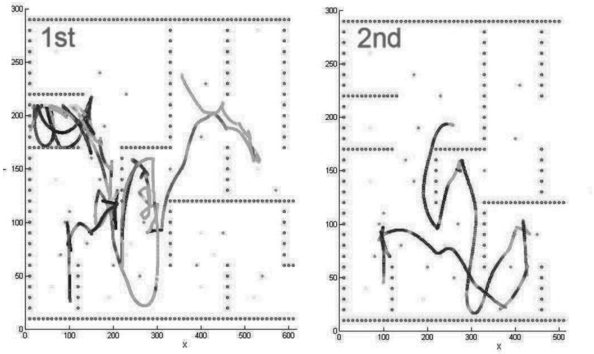


Figure 4: Trajectories performed by Group 1 during first and second Trials with change of velocity

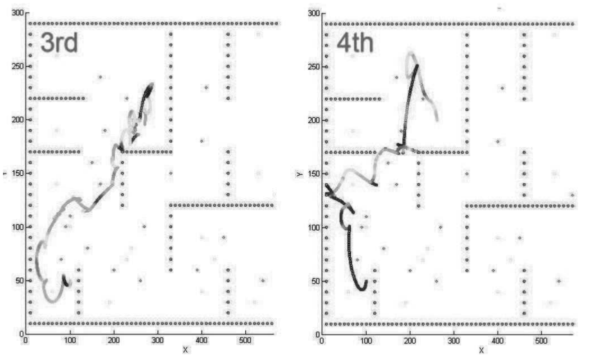


Figure 5: Trajectories performed by Group 1 during third and fourth Trials with change of velocity

at first could cause a difficulty during exploration and for this reason we decided to repropose the experiment four times with different visual knowledge so that the data collected wasn't affected by the initial interaction problem.

5 Results

A preliminary consideration about the results is that the few people with higher expertise on haptics got a better performance in terms of time of completion, cooperation with the other user and control of the navigation. We could aspect this result because all the interaction was mediated by the previously presented GRAB haptic interface. From the point of view of trials almost all groups had a progress from trail to trial except group 3 where one of the participant was very confused by the use of haptics.

From figures 4 and 5 is possible to see the trajectories performed by the first group in the trials with respect to change of velocity. At the first trial the trajectory is really confusing, in the second trial users were both able to see the entire maze producing a smoother trajectory. Finally they were been able to localize the goal. Indeed the third and fourth trials have trajectories direct toward the goal, showing that the users have understood where the goal is. Analyzing the retrieved data we can assert that groups have used different strategies about the collection of bonuses. Some of them tried to collect the bonuses avoiding bomb, others decided not to worry about stars but to reach the goal as soon as possible.

Relating the cognitive flow from one group to another, only group 4 have received precise information and have performed the tests at best obtaining top timing scores with respect to the other participants. The smoothness of the resulting position trajectories (figure 6) is a confirmation of this result. In addition group 4 has shown better performance in terms of smoothness of the forces, because of a more precise coordination. Figures 7, 8 and 9 show profiles of forces for one trial of the group number 4. In particular is shown the delta between forces applied by the users to the haptic interfaces. Figure 10 presents a plot of data collected from the first group on the second trial, comparing the graphs with the one of the fourth group (figure 7) is possible to see that the usage of devices wasn't smooth for the first group, the same is for others trials and groups showing again that the usage of the devices have influenced the performances. In terms of forces it is worth noticing that users had a minimal use of the third axis that was not involved in the interaction during the maze.

From figure 11 it is possible to see the time elapsed to reach the end for each group in each trial. The unit of time is step of simulation. Figure 12 shows bombs (maluses) collided by each group during the tests and figure 13 shows the stars (bonuses) collected.

About the questionnaire it resulted that user considered the task not easy to perform even if, in their opinion, they were satisfied with the support informations given (instruction, training) and considered to have completed the task properly. The results show the adopting of a strategy growing from trial to trial. Users also stated that the strategy changed along the trials.

6 Main Conclusions

This paper shows the preliminary results of an experiment for the identification of cooperative strategy between users through the haptic channel. From trail to trail users gained knowledge and cognitive information leading to a better cooperation between group partners that resulted in a more direct trajectory toward the end of the maze proposed. The creation of a strategy however was evident only in users that had previous haptic experience confirming that the channel of user interaction has a primary role in the experiments. More inves-

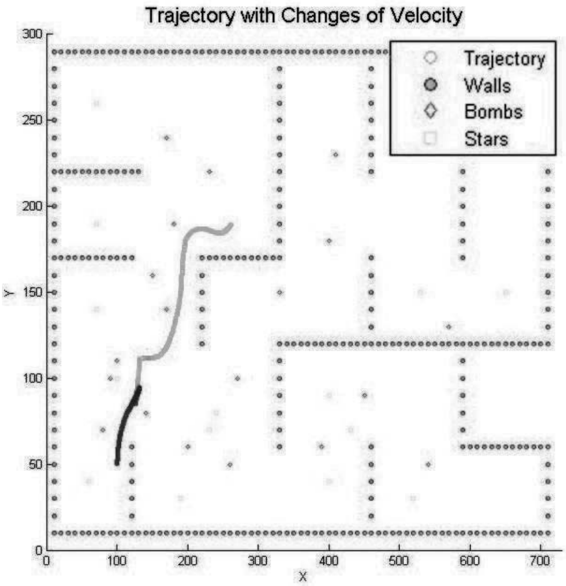


Figure 6: Trajectory of Group 4 in Trial 3

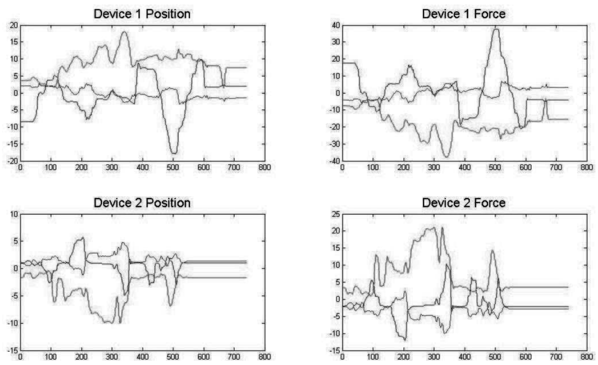


Figure 7: Device plots for Group 4 in Trial 3

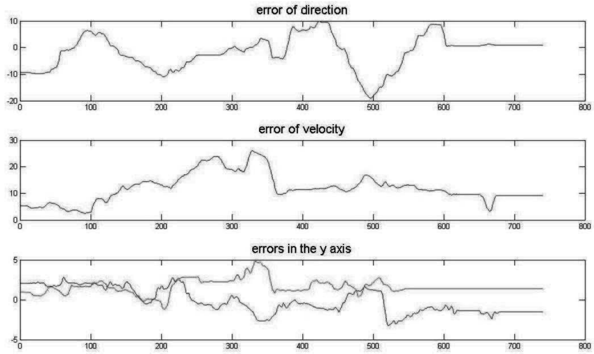


Figure 8: Errors between users for Group 4 in Trial 3

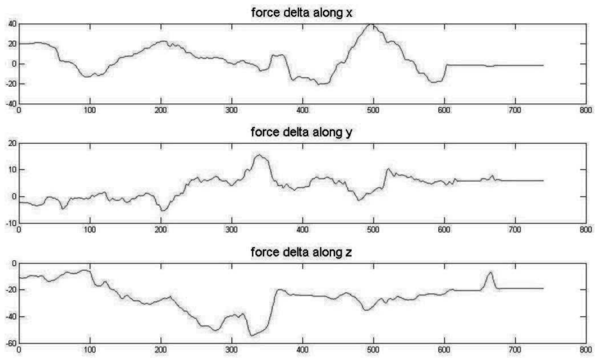


Figure 9: Force profiles for Group 4 in Trial 3

tigation is necessary to model the flow of information through the haptic channel.

7 Acknowledgments

This work has been carried out within the context of the Project IST-2004-002114- ENACTIVE Network of Excellence European Project and the Project IST-2006 SKILLS.

References

[1] R. I. Brafman and M. Tennenholtz. Modeling agents as qualitative decision makers. *Artif. Intell.*, 94(1-2):217–268, 1997.

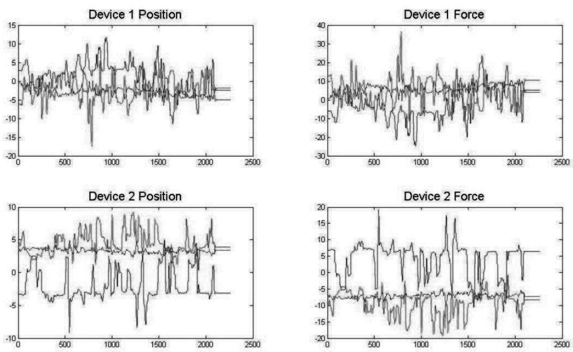


Figure 10: Device plots for Group 1 in Trial 2

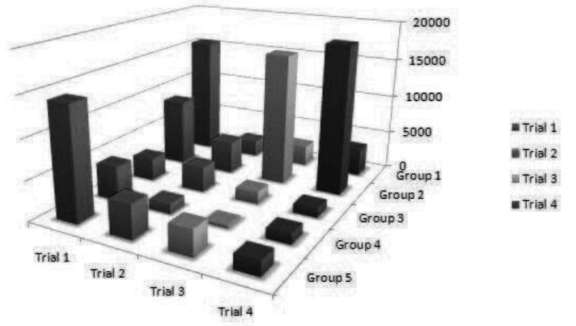


Figure 11: Time steps ellapsed during trials

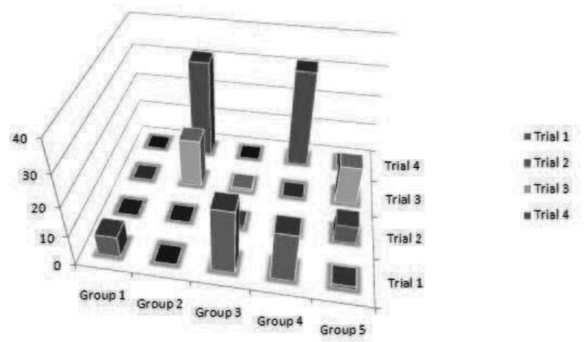


Figure 12: Bombs collided during trials

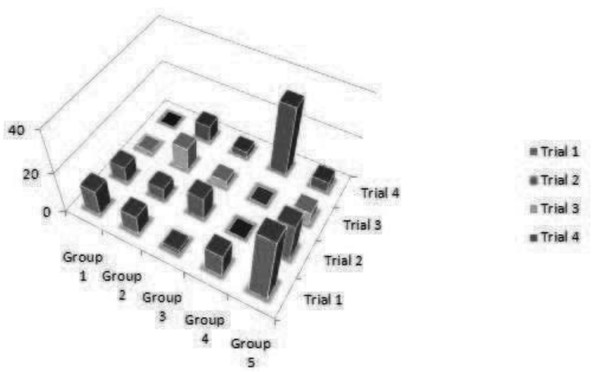


Figure 13: Stars collected during trials

[2] B. J. Grosz and S. Kraus. Collaborative plans for complex group action. *Artificial Intelligence*, 86(2):269–357, 1996.

[3] C. C. Hayes and C. C. Hayes. Agents in a nutshell-a very brief introduction. *IEEE Trans. on Knowl. and Data Eng.*, 11(1):127–132, 1999.

[4] R. Iglesias, S. Casado, T. Gutierrez, J. Barbero, C. Avizzano, S. Marcheschi, M. Bergamasco, F. Labein, and S. Derio. Computer graphics access for blind people through a haptic and audio virtual environment. In *Haptic, Audio and Visual Environments and Their Applications, 2004. HAVE 2004. Proceedings. The 3rd IEEE International Workshop on*, pages 13–18, 2004.

[5] S. Kraus. Negotiation and cooperation in multi-agent environments. *Artificial Intelligence*, 94(1-2):79–97, 1997.

[6] V. R. Lesser. Cooperative multiagent systems: A personal view of the state of the art. *Knowledge and Data Engineering*, 11(1):133–142, 1999.

[7] B. Shneiderman. Creating creativity: user interfaces for supporting innovation. *ACM Trans. Comput.-Hum. Interact.*, 7(1):114–138, 2000.

[8] M. Tambe, J. Adibi, Y. Al-Onaizan, A. Erdem, G. A. Kaminka, S. Marsella, I. Muslea, and M. Tallis. Using an explicit model of teamwork in robocup-97. In *RoboCup-97: Robot Soccer World Cup I*, pages 123–131, London, UK, 1998. Springer-Verlag.

[9] M. Tambe, J. Adibi, Y. Alonaizon, A. Erdem, G. A. Kaminka, S. Marsella, and I. Muslea. Building agent teams using an explicit teamwork model and learning. *Artificial Intelligence*, 110(2):215–239, 1999.

[10] P. Tripicchio, E. Ruffaldi, C. Avizzano, and M. Bergamasco. Virtual laboratory: a virtual distributed platform to share and perform experiments. *Haptic interfaces for virtual environment and teleoperator systems, 2008. haptics 2008. symposium on*, pages 311–318, March 2008.

[11] D. S. Wilson. Cognitive cooperation: When the going gets tough, think as a group. *Teamwork: Multi-Disciplinary Perspectives, N. Gold, Ed. New York, NY: Palgrave Macmillan*, pages 33–55, 2005.

[12] H. Zhuge. Conflict decision training through multi-space co-operation. *Decis. Support Syst.*, 29(2):111–123, 2000.

Information-action model for skills transfer

Emanuele Ruffaldi* Carlo Alberto Avizzano*† Massimo Bergamasco‡
(*) Scuola Superiore S.Anna, Pisa, Italy
E-mail: e.ruffaldi@sssup.it,carlo@sssup.it,bergamasco@sssup.it

Abstract

Design of Virtual Reality training systems has focused mostly on technological aspects and specific domains. The application of multimodal training to new areas stimulates the research on more general approaches of the skill transfer process and of the design of such trainers. This work proposes a information-action model that describes how the interaction between agent, environment and trainer can take place, in particular with the possibility of taking into account different solutions of virtual reality.

1. Introduction

Virtual environment and virtual reality technologies have found an interesting area of application for the training of human in a variety of tasks, from space [8] to medicine [12]. The improvements in multimodal feedback technologies, capturing systems and machine learning gave the opportunity for exploiting areas of research that have been typically limited to expensive virtual environment setups. The analysis of human motor skills, and the techniques for improving them, are aspects discussed in several virtual reality simulators although with the focus on specific domains [1].

A skills is a human ability that can be learn and it is expressed by specific characteristics ([13]): reduced mental load, repetiveness of performance, reduced attention requirements, reduced energy consumption and minimal errors [11]. A skill can be identified through specific signs, sometimes called the signature of skill [6], but most of the time it requires the interaction with experts. The acquisition of a skill is connected to the concept of Enactive learning, and in general it exploits the Enactive memory of humans, that is the memory of doing [4].

The theoretical research on human behavior has produced interesting works and experiments that anyway have received not sufficient interest by training system designer. This work is aimed at presenting a model that can be applied for the design of a virtual environment training system, with the specific focus on the transfer of skills. The aim of this model is to be applied to different domains while keeping the fundamental principles.

The paper is structured as follows: the next section contextualizes this research and the objectives of the modeling of skills. Then follows a description of the proposed model, in the form of evolution from existing approaches to a more flexible solution. Finally the fourth section applies the proposed model to two examples in working training systems.

2. Research Context

The design of virtual environments for training has mostly focused on technological improvements for achieving higher realism. There are anyway two other aspects of virtual reality systems that are extremely important for the effectiveness of the training, the performance evaluation and the training feedback.

The initial approach based on a simple assessment of the overal task performance has been superseded by an analysis of user movements and actions, allowing objective evaluations of quality of the task performance [12, 3]. Such analysis can be easily performed in tool mediated contextes like robotic surgery [14], or can be supported by high quality motion sensing technologies in free body contextes [2]. Because the objective of these systems is the transfer of the human skill from virtual to real environment, it is necessary to evaluate this transfer and understand how the multimodal technology is able to perform such transfer [7]. In [20] and [15] instead emerged the idea of using robotic systems for guiding the user for improving the learning process.

Models of human behavior have been extensively addressed in literature, but fewer propose a discussion on learning and training. The exchange of information between behavioral components has been introduced by [18], while [17] introduced a dynamic model that covers both the agent and the external environment. The learning process has been addressed by [16, 9, 21] with focus on behavioral elements.

This paper starts from the information based approaches presented above and from the control based approaches of many virtual environment designs into a model that is able to describe digital based training systems. The first claim of this work is to move beyond works totally based on the dynamics of physical variables to take into account reduced space coming from machine learning techniques. Moreover the proposed model takes into account both single and multiuser training and it is capable of describing real and virtual training. Finally the whole approach is based on the concept of Enactive learning.

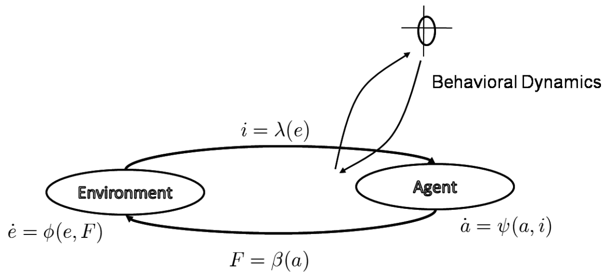


Figure 1: Warren model of dynamic perception action. This model describes the behavior of an Agent interacting with the Environment. The description is made in terms of dynamic states and flow of information.

3. A model for interaction and training

The analysis of interaction and training models presented above is the background for understanding the characteristics of the unified approach discussed in this paper. The aim of this section is to present a formal model of the training system and in general the interaction. The objective of this model is to identify the types of information and the roles of the entities present in the system.

The starting point of the model presented here is a dynamic perception-action model introduced by Warren [17]. In this model the behavior of a human, or of an agent in general, is modeled in terms of exchanges of information between the Agent and the Environment, each described as a dynamical system. The Environment has an internal state e that is mapped into the Agent through a function λ. This first mapping corresponds to the perception of the Agent, that depends on the perceptual system of the Agent itself. The information i generated by the mapping $\lambda(e)$ affects the Agent through an internal function ψ. The Agent is represented by an internal state a with a dynamic expressed as $\cdot a = \psi(a, i)$. The result of the change in the state of the Agent is the action F. The action F is obtained through a mapping function $F = \beta(a)$ and changes the Environment. Indeed we can express the dynamic of the environment as $\cdot e = \phi(e, F)$. The result of this model is an interesting cyclical flow of information and changes of states, that can be represented as in figure 1.

The model of Environment-Agent interaction allows to describe the behavior of the Agent in a dynamic Environment, but is not able to capture directly the aspects of training and learning that are the focus of our research.

The first variation to this model is the case of multiple Agents interacting in the environment: in this case information flows between each Agent and the Environment, that is envisioned as a mediator of the action and perception of the users. The introduction of additional Agents affects most of the mapping functions in the model, in particular the internal strategy of the Agent ψ and the dynamics of the Environment. In addition, the perceptual mapping λ changes for many reasons, and also when the Agent has to perform a task without any collaborative effort, the presence of other Agents changes its attention, consequently affecting the mapping.

The multi-agent model supports the introduction of the interaction between Trainer and Trainee, as shown in 2. The considerations about the Training methodologies and the feedback of the Trainer to the Trainee, can be expressed by this model. For this occasion we extend the notation using the super script for specifying the Agent to which a given function or information is relative. The first consideration is relative to the states. In particular the Trainee-Agent A is affected by the presence of the Trainer both in terms of perceptual mapping λ^A and internal state function ψ^A. The state of the Trainer-Agent T corresponds to the understanding of the state of the Trainee and his behavior, and we can associate the ψ^T function to the evaluation process performed by the Trainer. Given this state the Trainer decides how to affect the Environment for improving the learning process of the Trainee; in particular it is possible to distinguish the β^T in terms of general feedback to the Environment $\beta^T_E(s)$ and specific actions toward the Trainer $\beta^T_A(s)$, like vocal commands. The perception of the behavior of the Trainee from the point of view of the Trainer is mediated by the environment, and in general it is not direct, as expressed by $\lambda^T(e)$. The introduction of biometric sensors allows the Trainer to have a direct connection to the Trainee-Agent, improving the knowledge of the state a of the Trainee itself.

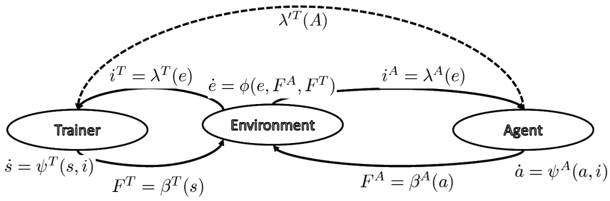


Figure 2: Model of information flow extened with the Trainer-Trainee interaction. In particular we are interested in the meaning and in the transformation of the mapping functions. The dashed line corresponds to the introduction of biometric information for having a direct knowledge of the state of the Agent.

3.1 From Real to Virtual Environments

A physical Trainer has a limited capacity in the types of feedback that it can provide to the Agent, and in addition it has a reduced view of the environment itself. These limitations can be overcome by the introduction of a Digital Trainer system. This system has a structure similar to the Trainer-Trainee with some fundamental differencies. First we can say that the space of $\lambda_E^{DT}(e)$ is larger than $\lambda_E^T(e)$ because the Digital Trainer is able to capture more information from the Environment. The second aspect is the nature of the Environment itself. In this system the Environment is made by two entities, the Real Environment, and the Virtual Environment. This distinction is important because the Digital Trainer is able to modify parameters of the Virtual Environment, like gravity or other physical properties. The original β^T becomes a more complex β^{DT} that allows to exploit new types of feedback during training. The β^{DT} can be described in terms of four components: β_E^{DT} are the actions toward the environment, β_A^{DT} are the actions sent directly to the user like vocal commands, or biofeedback, β_I^{DT} is the information sent to the user and finally β_P^{DT} represent the capacity of changing the Environment variables. Given these four types of action, and the capability of the Digital Trainer it is possible to understand the main differences respect the human Trainer. In particular the DT is able to act more effectively into the Environment than the Trainer, that is $\beta_E^{DT}(s) > \beta_E^T(s)$, and at the same time the DT is able to produce multimodal feedbacks to the Agent, that is $\beta_A^{DT}(s) > \beta_E^T(s)$. The overall scheme of the Digital Trainer with the explicit representation of the four types of feedback is shown in Figure 3.

The distinction of the Environment in two entities, Real and Virtual, gets complicated when we take into consideration the whole Mixed Reality spectrum as defined in [10]. While it is important to keep the general scheme of inputs and outputs of the Environment as a whole it is necessary at the same time to identify the characteristics and the interactions between the components of this generalized Environment. In a purely Virtual Reality setup the Real environment should be modeled only for taking into account the effects of illusions and problems of sickness, while in an Augmente Reality solution the entities and the $\beta_E^{DT}(s)$ should take into account both the real environment and the virtual entities generated on the perceptual field of the Agent.

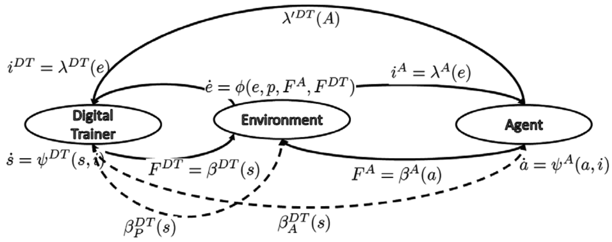


Figure 3: Model of the Digital Trainer showing the various type of feedback. In particular it is important to understand how the Trainer is capable of changing the physical parameters of the Virtual Environment.

3.2 Enactive Digital Trainer

The Digital Trainer, based on the integration of Virtual and Real Environment, is quite generic, and it important to clarify how the training mechanism works and how the technology affects the interaction with the Agent. In this paper we are introducing the concept of the Enactive Digital Trainer that is a specific type of Digital Trainer based on the concept of Enactive Interfaces. The main property of the Enactive Digital Trainer is that it is based on Learning by Doing, in which the Agent learns directly from the interaction with the Environment and not from commands from the Trainer. In particular we can characterize the Enactive Digital Trainer from its focus on the transformation of the environment, in particular on the capacity of affecting the basic parameters of the Environment $\beta_P^{DT}(s)$. In the case of the Enactive Digital Trainer the informative feedback $\beta_I^{DT}(s)$ is reduced at most, being replaced by transformations in the environment that directly modify the Agent behavior [19].

3.3 Inside the Digital Trainer

After the description of the information exchange among the components of this model it is necessary to describe the structure of the Digital Trainer. The function $\psi^{DT}(s, i)$ has the role of performance evaluation based on the inputs coming from the environment and the agent. Such evaluation is based on many factors and it can be structured using the sub-skill description of the skill is being trained. The factors affecting the evaluation functions are Agent history, style and the adopted skill model based on experts' knowledge. The feedback generated by the Digital Trainer is represented by the function β^{DT} that takes into account the knowledge of experts for guiding the Agent, basic rules coming from a task analysis [5] and most importantly the accelerators available to the training system. An accelerator is a specific type of feedback, that can be activated for improving the training process. For example in a juggling demonstrator the possibility of reducing gravity or modifying in general the space around the user is some form of accelerator.

Finally the overall scheme, involving expert knowledge and the training dynamics, is shown in Figure 4.

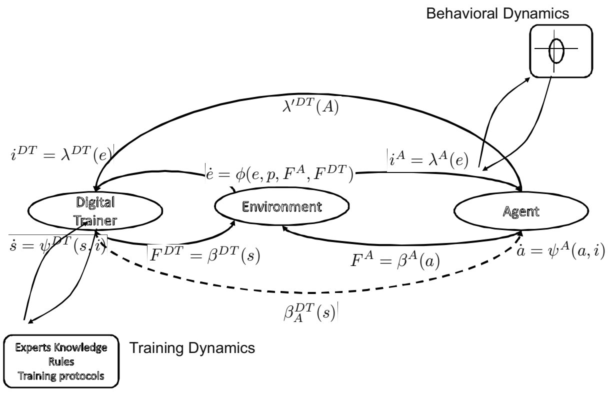


Figure 4: Complete scheme showing the relationships between the Environment, the Agent and the Digital Trainer. Specific emphasis is given to the connections between the training system and the experts' knowledge that is used for driving the training dynamics.

3.4 Multi user Training

The model presented above can be easily extended to the case of tasks in which multiple users are involved, while at the moment the focus is on the training of a single Agent. In the digital trainer the other Agents can be virtual, allowing the training system a new degree of flexibility in the training process. Every Virtual Agent has a behavioral model that affects the Virtual Environment with its own independent feedback. Anyway the behavior of these virtual agents can be modified by the Digital Trainer using specific actions for indirectly affecting the target Agent. An example of such Virtual Agents is the case of training for team rowing in which only one Agent is real: the other team members are simulated with behaviors that are controlled by the digital trainer for maximizing the effectiveness of the training session.

4. Using the model

The model described above can be better understood with some examples of its use, and how the model is mapped into a real working system. Before introducing these examples the model should be put into context of the complete skills analysis and transfer pipeline adopted. This pipeline describes the way the skill is analyzed for transfer, represented and then transferred to the user. There are four steps in the pipeline:

1. modeling of the skill - creating a high level decomposition of the skill in sub-skills and identifying the major characteristics;

2. capturing of the skill - using capturing technologies for extracting skilled behavior from experts;

3. encoding of the skill - use of the captured data and additional knowledge to digitally encode the skill;

4. transferring the skill - use the encoded skill for transferring it into another person.

The first three steps of the pipeline can be applied multiple times, while refining the encoding of a skill, and also it can be applied at a lower level in each component of a given skill. In the transfer phase of the pipeline, the target user performs a given task and in real time its behavior is represented in the skill space for identifying the performance level and the appropriate feedback for correcting its actions.

In the design phase of a trainer it is possible to analyze the task objective for completing the various elements of the model described above. In the first phase the task is decomposed in terms of sub-components and the same operation is performed with the skills, decomposing them in sub-skills. This operation is required for identifying the sub-domains of the problem that should be addressed by the applied model. From the decomposition it is possible to proceed to the performance indicators, taking also into account information from experts. These two phases are used to build the state based description of the Environment. In the final phase of the instantiation of the model, the behavior of the digital trainer is selected depending on the chosen training method.

This section continues with two examples of application of this model, in particular a juggling trainer based on enactive learning and a rowing trainer.

4.1 Juggling Enactive trainer

The juggling Enactive trainer is platform for performing experiments on the training of juggling that is organized around the model described in this paper. In this platform the user stays in front of a projection screen with its hand tracked and performs some juggling exercises. The feedback is mostly visual, although some preliminary solutions of vibrational feedback have been introduced.

In this platform the Environment state e corresponds to the position and velocities of the balls while the parameter p is associated to the gravity vector and the air friction that can be used for affecting the training. User's hands participate as F_A to the environment in terms of environment's constraints , while the movements of the hands are not needed for measuring the performance of the user, but only for measuring the quality of the movements itself. Figure 5 shows the description of the Environment in this scenario.

In terms of Digital Trainer we can identify the following characteristics. The input from the environment i^{DT} contains the status of the balls and descriptors of the hands' movements. The state s of the Digital Trainer is expressed in terms of sub-skill characteristics, like the synchronization of the bi-manual coordination, and the overall performance index. This state is updated by $\psi^{DT}(s,i)$ using the sensing information i. The feedback generated from s using $\beta^{DT}(s)$ involves direct effects over the balls and changes in the environment parameters. The extension of the feedback with vibrotactile stimulation is feasible in this model, introducing a $\beta_A^{DT}(s)$ that goes directly to the Agent.

The model of the user in the environment is not directly user by this system, although it can be used for taking into account perceptual effects of the virtual environment as of $i^A = \lambda^A(e)$.

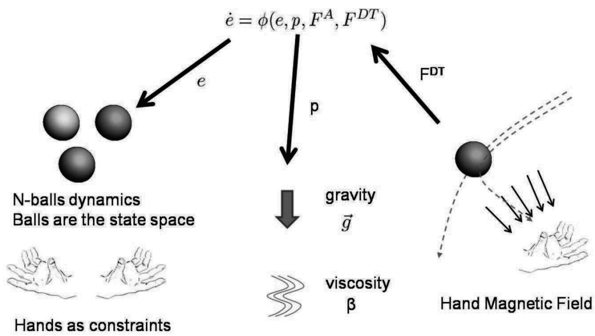


Figure 5: Description of the Environment in the juggling scenario, showing the basic state, the parameters and the feedback from the trainer

4.2 Enactive rower

The other example of the application of this model is a training system for rowers, that uses motion capture, visualization and haptic feedback for improving rower's performance. Figure 6 presents the usual model scheme with highlights on the specific elements. In particular the environment state takes into account the boat, the landscape scenario and the oars. The training feedback is expressed by the fan regulator and the visual display. In addition the Digital Trainer has been trained using recordings and rules from experts, affecting the performance indices and the evaluation function.

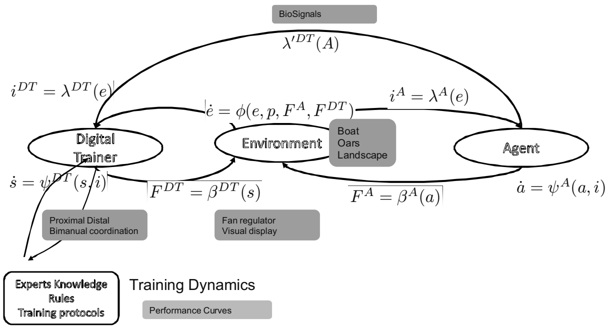


Figure 6: Description of the Environment in the rowing scenario, showing the entities that define states of the environment and exchanged information

5. Main Conclusions

This paper introduces a proposal for the modeling and the design of training system capable of taking into account both real and virtual environments. This model can be effectively implemented with multimodal technologies and can be applied to different types of training methods. The research in this model continues in the area of valdiation and in the specific adaptations and cases in various mixed reality contextes.

6. Acknowledgments

This research has been performed inside the European Community funded project SKILLS (FP6-IST-2005-035005-IP). The authors thank the EU for supporting this research.

References

[1] R. Adams, D. Klowden, and B. Hannaford. Virtual Training for a Manual Assembly Task. *Haptics-e*, 2(2):1, 2001.

[2] P. Chua, R. Crivella, B. Daly, N. Hu, R. Schaaf, D. Ventura, T. Camill, J. Hodgins, and R. Pausch. Training for physical tasks in virtual environments: Tai Chi. In *Proceedings of IEEE Virtual Reality*, pages 87–94, 2003.

[3] A. Darzi, S. Smith, and N. Taffinder. Assessing operative skill: Needs to become more objective. *BMJ: British Medical Journal*, 318(7188):887, 1999.

[4] M. Foley, V. Bouffard, T. Raag, and M. DiSanto-Rose. The effects of enactive encoding, type of movement, and imagined perspective on memory of dance. *Psychological Research*, 53(3):251–259, 1991.

[5] J. Frederiksen and B. White. An approach to training based upon principled task decomposition. *Acta Psychologica*, 71(1-3):89–146, 1989.

[6] A. Hurst, S. Hudson, and J. Mankoff. Dynamic detection of novice vs. skilled use without a task model. *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 271–280, 2007.

[7] K. Lehmann, J. Ritz, H. Maass, H. Çakmak, U. Kuehnapfel, C. Germer, G. Bretthauer, and H. Buhr. A Prospective Randomized Study to Test the Transfer of Basic Psychomotor Skills From Virtual Reality to Physical Reality in a Comparable Training Setting. *Annals of Surgery*, 241(3):442, 2005.

[8] R. Loftin and P. Kenney. Training the Hubble space telescope flight team. *Computer Graphics and Applications, IEEE*, 15(5):31–37, 1995.

[9] D. Meyer and D. Kieras. A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychol Rev*, 104(1):3–65, 1997.

[10] P. Milgram and F. Kishino. A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information Systems*, 77(12):1321–1329, 1994.

[11] S. Mitra, P. Amazeen, and M. Turvey. Intermediate motor learning as decreasing active (dynamical) degrees of freedom. *Human Movement Science*, 17(1):17–65, 1998.

[12] K. Narazaki, D. Oleynikov, and N. Stergiou. Robotic surgery training and performance. *Surgical Endoscopy*, 20(1):96–103, 2006.

[13] M. Posner. Components of Skilled Performance. *Science*, 152(3730):1712–1718, 1966.

[14] J. Rosen, J. Brown, M. Barreca, L. Chang, B. Hannaford, and M. Sinanan. The Blue DRAGON-A System for Monitoring the Kinematics and the Dynamics of Endoscopic Tools in Minimally Invasive Surgery for Objective Laparoscopic Skill Assessment. *Studies in Health Technology and Informatics-Medicine Meets Virtual Reality*, 85:412–418, 2002.

[15] A. Sano, H. Fujimoto, and K. Matsushita. Machine mediated training based on coaching. In *Proceedings of IEEE Systems, Man, and Cybernetics SMC'99*, volume 2, 1999.

[16] R. Schmidt and T. Lee. *Motor Control And Learning: A Behavioral Emphasis*. Human Kinetics, 2005.

[17] W. Warren. The dynamics of perception and action. *Psychological Review*, 113(2):358–389, 2006.

[18] C. Wickens and A. Kramer. Engineering Psychology. *Annual Reviews in Psychology*, 36(1):307–348, 1985.

[19] G. Wulf and C. Weigelt. Instructions about physical principles in learning a complex motor skill: to tell or not to tell.... *Res Q Exerc Sport*, 68(4):362–7, 1997.

[20] Y. Yokokohji, R. Hollis, T. Kanade, K. Henmi, and T. Yoshikawa. Toward machine mediated training of motor skills. Skill transfer from human to human via virtual environment. In *5th IEEE International Workshop on Robot and Human Communication*, pages 32–37, 1996.

[21] P. Zanone and J. Kelso. Coordination dynamics of learning and transfer: collective and component levels. *J Exp Psychol Hum Percept Perform*, 23(5):1454–80, 1997.

Posters

Requirement for an Enactive Machine : Ontogenesis, Interaction and Human in the Loop

Pierre De Loor* Kristen Manac'h* Alexandra Fronville* Jacques Tisseau*
(*) Université Européenne de Bretagne - ENIB - LISyC
Centre Européen de Réalité Virtuelle 25, rue Claude Chappe, F-29280 Plouzané (France)
E-mail: deloor@enib.fr, manach@enib.fr, alexandra.fronville@univ-brest.fr, tisseau@enib.fr

Abstract

This article deals with the links between the enaction paradigm and artificial intelligence. Enaction is considered as a base for some artificial life and robotic approaches. We explain some technical and conceptual problems relative to the development of such approaches: Lack of complex ontogenetic mechanisms, and of understanding the notion of sense-making in an interactive artificial intelligence context. Then, we propose that 1) the growing complexity of the ontogenetic mechanisms to be activated can be compensated by an interactive guidance system emanating from the environment. 2) The integration of human into this environment leads to construct relevant meaning in terms of participative artificial intelligence. This raises a number of questions with regards to setting up an enactive interaction. The phenomenology of interactions and the use of minimal enactive interfaces in setting up experiments will deal with the problem of artificial intelligence in a variety of enaction-based ways.

1. Introduction

Enaction and its revolutionary vision of cognition enable to lay down new foundations for artificial intelligence. These new foundations led [8] to lay down the guidelines for *Enactive Artificial Intelligence*. We will remain prudent about the terms we use, considering enaction as a metaphor for artificial intelligence. We shall therefore refer instead to Enaction-Based Artifical Intelligence (EBAI). Indeed, the direct transfer of a paradigm from the cognitive sciences might lead to shortcuts, misunderstanding and confusion regarding the initial notions of the paradigm. For example, enaction borrows the specificity of first-hand experience from phenomenology, and it is necessary to use phenomenology in order to understand the mind. However, in the case of machines, the notions of first-hand experience, consciousness and own-world are without a doubt inaccessible, if not absurd. We shall simply embark on an analysis of [16] who propose to replace the traditional distinction between ascriptional and genuine autonomy by presenting the hypothesis that an attributional judgement based on knowledge of an underlying behavior-inducing mechanism will be more stable than a naïve judgement based only on observation of behavior. This concept enables us to use the ideas and advances of cognitive science in order to contribute to the artificial sciences and vice versa. In particular, for [19], artificial dynamical approaches provide an alternative to the frame problem. [18] argue that the "Rosen's conjecture" - i.e. that a model of life defined as closure under efficient cause cannot be instanciated by a computer simulation - is false. Furthermore, [4] argues that the problem of sense-making, crucial in artificial intelligence, might be established in an enactive inspiration.

This article tends to clarify some directions which must be followed to go beyond the bound of actual approaches. In particular, we will argue the necessity to better include the environment implication during the developement of the ontogenesis of an artificial model, to pass from an phylogenetic point of view (used in evolutionnary robotic) to an ontogenetic one and to integrate human-machine interaction at the basis of a human-like sense-making.

This article will be structured in the following manner: section 2 point out the main elements of what should be an enactive machine. We will show how little importance is given to the evolution of the environment by current approaches. The notion of sense-making for a machine can also be a problem for a human user if it is designed to be autonomous in a purely virtual world. Having studied these issues, we make a number of suggestions in section 3: a more explicit recognition of the irreversible evolution of the environment and of coupling; guiding the artificial entity in order to tackle more complex ontogeneses as is the case in the co-evolving nature and integration of the ”man-in-the-loop” with the co-creation of meaning, compatible with the social construction of meaning and the initial precepts of AI, illustrated by the TURING test. The section 4 then goes on to present the areas which we shall explore in future research in order to meet these goals, before going on to the conclusion (section 4).

2. Enactive machine ?

2.1 Requirement

Enaction proposes to address cognition as the history of structural coupling between an organism and its environment. It originates from the notion of autopoietic systems put forward by MATURANA and VARELA as model of the living centred on the capacity of organisms to preserve their viablility [24]. For these authors, this preservation defines the organism's autonomy and constitutes the biological origin of its cognitive capacities. An autopoietic system is a structure which produces itself as a result of its environment. The environment may disrupt the system, whose functioning will evolve as a consequence of that effect. If the functioning of the organism evolves in such a way as to preserve it despite disruption from exterior factors, the organism can be considered viable. This new way of functioning will, in return, influence the environment and the organism-environment system will co-evolve. The fact that the environment is but a disruption implies that it does not seem to be represented within the organism as a pre-given world. Furthermore, constraints on viability and the necessity to remain alive endow the organism with an identity by means of its metabolism and its capacity to act. This identity emerges relative to viability constraints, and the environment gradually takes on meaning. It is worth noting that co-evolution is usually considered at a phylogenetic scale, but, in an enactive perspective, it can be also considered at an ontogenetic scale. The example is often giving of tracing a path by trampling the ground with our feet. Another favorite topic in enaction is those of the constitution of knowledge in the form of sensorimotor skills by continuous back and forth between actions and perceptions [14]. Thus, as MERLEAU-PONTY notes, the organism both initiates and is shaped by the environment. For more details on the perspective of enactive cognitive science, we recommend the review articles by [12]

Breaking away from biology, we talk about operationally closed systems. An operationally closed system forms a system of recursively interdependent processes in order to regenerate themselves, and can be identified as a recognizable unit in the domain of processes. One important question is to know if nothing prevents to apply the notion of operationally closed systems to the phenomenal domain of the artificial. [18] well illustrate the debate on the possibility of such a simulation, argued about the ROSEN objections. Nevertheless, the biological argumentation would then be to generalize this mechanism to multi-cellular organisms [23], and thus to human beings, the mind, and social cognition [3]. At each level, there is a difference linked to the aspects associated with the notions of viability and unity. Without entering into further detail and the arguments behind the theoretical approach, we shall retain three important characteristics involved in the development of artifical systems based on this paradigm:

1. Operational closure and dynamicity which lead to an absence of a priori representations: In the domain of AI, this conclusion shares similarities with R. BROOKS considerations about the postulate that intelligence doesn't need representation [2].
2. Plasticity: It can be observed not only in the body for physical interactions but also at nerve level for higher-level interactions (cerebral plasticity).
3. Co-evolution: At the ontogenesis scale, it means that interactions between the system and its environment can produce modifications of each one. The story of these interactions progressively shapes the structure, the dynamic and the behavior of the system. And this story leads toward a result which is irreversible.

In this way, we can see that the artificial system is taking the form of a complex system i.e. it is heterogeneous, with an open and multi-scaled dynamic. The emergent properties of these systems are testimony to the openness and the multiplicity of the possibilities of evolution. The notion of ”natural derivation”, highly important in enaction is thus converted to ”artificial derivation”. It underlies complex systems and can initiate creativity and commitment in ”bringing forth a new world”. Creativity is here defined as the possibility to determine the functions of an undefined element of the environment.

Co-evolution involves a recursive transformation of the system and of its environment. The environment is thus an actor in the same way as the entity that occupies it.

2.2 Artificial Autopoiesis and Enactive Robotics

Not surprising, artificial life and robotic are the field which give rise to an enactive inspiration. For artificial life, the challenge is to set up autonomous models lying on autopoiesis concept. Such systems are compounded of elements which are generated by the system itself. Generally, simulations lie on cellular automata [13] but this proposition is very limitative. Indeed, it confines the concept of viability to the persistence of a topological aspect of the automata. However, in the numerical domain, such topology is far to be significant. There are some other ways to obtain self-organized models such as it is achieved in the multiagent research field. Furthermore, it is interesting to question the influence of the choice of an abstract level to implement the operational closure concept for an applicative purpose. At last, autopoiesis simulation neither adress the notion of co-evolution nor such of morphogenesis or interactive machine, coupled with a human user or with other autopoietic systems. Existing works insist on the autonomy which is obviously fundamental to clarify in the enactive perspective, but which isn't sufficient in itself, for a human-machine interaction perspective. In robotic, works linked with an enactive view, adress the problem of the self-organisation of a dynamical system [5, 20]. The aim is to find principles which can lead to the setting up of sensorimotor invariant. Basically, dynamical models such as recurrent neural networks are used because they approximate complex systems and they maintain dynamics without needs of something like input control. Environment is not more than an element that disturbs this dynamics. Ultrastability principles and evolutionary approaches are generally brought together to find 'off-line' plasticity rules which lead to an auto-adaptive system. Amazing results are thus obtain which refer to psychological experimentation on human sensorimotor adaptation or learning [9]. But the evolutionary approaches are faced with the problem of phylogenetic/ontogenetic articulation, which seems to be extremely difficult to resolve: Genetic algorithms must evaluate individual performance of each individuals (in order to select and improve the best). Idea of performance involves the pre-definition of a task to be performed. And so, if the task is determinated, then the adaptation remains focused on one final behavior. If we want to obtain co-evolution with some caracteristic of irreversibility, we must try to inverse the ratio between ontogenesis and phylogenesis.

3 Proposition

We shall now go on to present a proposition that aims to push back the limits previously identified here so as to enable an EBAI to refine its agentivity by means of more complex co-evolution. This proposal is based on the following arguments involving irreversibility, ontogenesis and sense-making.

3.1 The problem of irreversibility

The irreversibility of co-evolution is often overlooked. It is also the case of the evolution of the environment, which follows the actions of the agent. All these elements are neglected in favor of initiating an adaptivity to external changes, i.e. those which do not follow the actions of the agent itself. We suggest that the agent should actively modify an environment which, in turn, should also evolve. This principle is based on research suggesting that an entity's environment is made up of other similar entities [6]. In the following section we shall present our arguments to support the hypothesis that this is not sufficient to control this co-evolution nor to enable it to access sense-making which might be relevant to humans. First, we shall try to complete the principles suggested by [8] for the constitution of an agent from an enactive perspective, by a ”principe of irreversibility”.

- **EBAI irreversibility design principle :** an artificial agent must have the ability to actively regulate its structural coupling, depending on its viability constraints, with an environment which it modifies and for which certain modifications are irreversible.

This implies that it is possible that, as a result of an action, the agent's perception of its environment may be altered in such a way that it will never again perceive that environment in the same way. The fact that this only involves certain modifications and not all of them thus enables the agent to stabilize its coupling, which cannot be done in an environment which is too flexible. One difficulty is thus to find the balance between sufficient resistance for it to be able to remember the interactions, an ”en habitus deposition” [10], and sufficient plasticity for it to be able to evolve.

3.2 The problem of ontogenesis

Even if current models of autopoiesis and enactive robotics are complex in the sense that we call upon the notion of emergence in order to characterize their general behavior, their ontogenesis can be considered relatively simple. Either the principles of autopoiesis and stability are the sole focus of attention, to the detriment of the evolution of these principles or, the ontogenesis of the agent is defined using an evolutionary approach. However, the Darwinian inspiration behind the evolutionary approach is not compatible with an explanation of ontogenesis as it evaluates a whole agent. The agent is ready to function and fulfil the task that it has been selected for. That being said, if we want to progress in terms of capacity, and to broaden the cognitive domain of artificial agents, we must take into account the fact that the more complex agents are, the greater the ontogenetic component of their behavior is, compared to the phylogenetic component. Furthermore, as they develop, the influence of the environment becomes superior to the influence of genetic predetermination [15]. From an enactive perspective, evolution is considered more as a process of auto-organization than a process of adaptation. It is therefore important to distinguish between an auto-adaptive system and a system which learns [7]. For example, in robotics, it is necessary to express evolutionary research differently so that it does not rely on the selection of agents capable of fulfilling a task or of adapting to a changing environment, but rather on a selection of agents capable of ”adapting their adaptation” to that of the other and thus cope with new environments. This is debatable, as we could argue that the behavioral creativity of natural organisms is inherited from the adaptation characteristics selected throughout their phylogenesis. It remains nonetheless true that every organism's past conditions both its identity and what it will become, and especially so in the case of organisms with highly developed cognitive abilities [15]. Even if the aforementioned research shows that the principle of ultra-stability supports this argument, one important issue still needs to be addressed: that of the generalization of ontogenetic development principles. This problem is so tricky that we suggest associating evolutionary approaches with *guided online learning*, during ontogenesis. Here, we fall under a Vygotskian perspective according to which training constitutes a systematic enterprise which fundamentally restructures all of the behavioral functions; it can be defined as the artificial control of the natural development process [25].

3.3 The problem of sense-making

Let us imagine that the previous step has been achieved and that we know how to obtain an artificial system capable of co-evolution. Let us also imagine that we could imitate the environment of such a system in the same way as the system itself. There would be a co-evolution of these two entities. Both systems could engage themselves along ”uncontrollable natural derivations”. Enaction considers that a subject's world is simply the result of its actions on its senses. Thus, the presence of sensorimotor invariants evolving at the heart of an artificial system is the machine's equivalent of ”virtual sense-making” in the virtual own-world. What would this sense-making represent for an artificial system co-evolving with another artificial system? We must be wary of anthropomorphism, which is inappropriate here as the construction of meaning and sense for such machines cannot be compared to those of humans. We argue that meaning, coherent within the perspective of Man using the machine, and evolving from the co-operation between Man and machine, can only emerge through interactions with a human observer. This by no means leads us to question the value of experiments in evolutionary robotics for the understanding of fundemental cognitive principles, but rather to attempt to address the problem of sense-making. We must nevertheless take precautions, keeping in mind the potential impossibility of attaining such knowledge, just as [16] argue for the notion of autonomy. We simply wish to explore the leads which might enable us to come closer to one of the aims of artificial intelligence: the confrontation of a human user and a machine [22]. We hypothesise that, from an enactive perspective, one relevant approach would be to explore the sensorimotor confrontation between Man and machine. In this context, we believe that Man must feel the ”presence” of the machine which expresses itself by a sensorimotor resistance in order to construct meaning about itself. This idea of a presence, much like the TURING test, evaluates itself subjectively. This has notably been studied in the domain of virtual reality [1, 17] and enables us to link phenomenology and Enaction-Based Artificial Intelligence. [3] comment on the participatory aspect and on coordination as a basis for the construction of meaning in an enactive perspective. They say that *Meaning belongs to the relational domain established between environment and agent*. The actions of the other are as important as the actions of a subject in contributing to the enaction of its knowledge. Thus, we argue that the human's participation in this co-evolution will enable to create meaning relevant for him. If Man is not part of this loop, there is no intelligence. Inversely, with his participation, the coupling causes a own-world to emerge for the user. This raises the issue of the mode of interaction between Man and machine.

3.4 Summary of our proposals

To clarify our remarks, our proposals are summerized in the following paragraph:

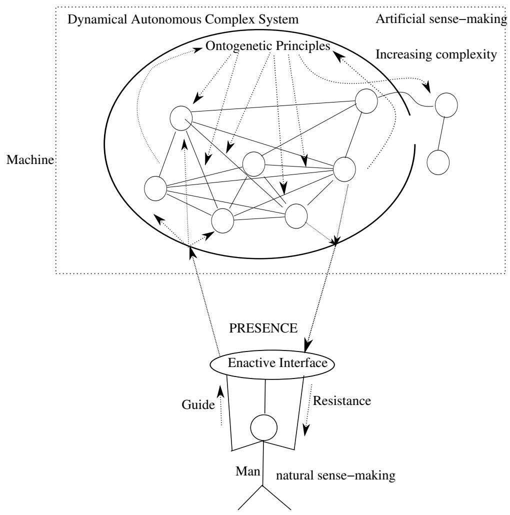


Figure 1: Artificial entity based on enaction metaphor

- **Proposal 1:** *To overcome the problem of irreversibility, we propose to add a principle obliging the agent to actively modify an environment which would also be evolving.*

- **Proposal 2:** *In order to overcome the issue of the complexity of ontogenesis, we propose the introduction of interactive guidance for the agent throughout its ontogenesis so as to leave it a memory of its interactions, as in the case of complex cognition in the animal kingdom.*

- **Proposal 3:** *To overcome the problem of the creation of relevant meaning in terms of the presence test, we suggest integrating humans into the loop so that a co-creation of meaning relevant to Man might correspond to relevant dynamics in the artificial system.*

These three proposals should not be addressed head-on. To us, it would seem appropriate to address the evolution of the environment without considering Man's presence in the loop or even to set up interactive guidance without addressing the environment. However, for each of these stages, we must not lose sight of the ultimate necessity for these two elements in order to guide the theoretical or technical choices that must be made when designing them. The final objective is to design ontogenetic mechanisms for complex dynamical systems which will be guided by people. This objective is illustrated in figure 1. Artificial entities are complex systems enriched with ontogenetic mechanisms which guide their evolution via an "en habitus deposition" of their interactions. This guidance can be conducted via a simulated environment, but must include human interaction. We think that it must be done using enactive interfaces. The complexity of online guidance such as this leads us to imagine progressive exercises linking the evolutionary and ontogenetic approaches. Actually, the work of [21], which guide physically his robots in order for them to learn sensorimotor skills and to progressively increase this skill, seems to be the more close to our objective. The main difference is that we want to use an enactive interaction perspective with artificial worlds in order to contribute to the understanding of the co-emergence of the sense during an interaction.

4 Perspectives

These premilinary reflexions lead us to sketch some interresting perspectives in the context of enactive interaction and virtual reality: Considerating the phenomenology of interactions between Man and machine in the constitution of sensorimotor skills for humans could be an important basis for establishing analogical principles for machines to tackle the problem of the inversion of the ontogenesis/phylogenesis ratio. Haptic interaction and virtual environment constitute a good base to develop guided models capable of co-evolution. However, we must remain prudent because of the incommensurable distance between the continous nature of the physical world which lead to the biological metabolism and the discret nature of numerical systems. Numerical and natural worlds are based on two different phenomenal domains and the later is tremendously more complex that the former. Nevertheless, it doesn't prevent the possibility to bring forth a world *into* a dynamical simulation, even if this world will be incommensurable with such of the human. The only interest for this artifical world would be in the fact that it would be constituted by the way of human-machine interaction and consequently that human might find a sense in these interactions. If it is the case, a man-machine common sense might be co-constituted. To do that, we must imagine experimentations easy enough to be supported by actual artificial models but also representative for a human in term of co-constitutive interaction. Artistic creation seems to be favorable to following this way. Technically, it is important to develop researches on ontogenesis and morphogenesis aspect of dynamical systems in order to avoid or to minimize the use of evolutionary approaches. We have done such an experiment in this direction in [11] in arguing that it is possible to obtain a continuous recurrent neural network which evolves by interactive guidance, during its ontogenesis.

References

[1] M. Auvray, S. Hanneton, C. Lenay, and J. O Regan. There is something out there: distal attribution in sensory substitution, twenty years later. *Journal of Integrative Neuroscience*, 4:505–521, 2005.

[2] R. Brooks. Intelligence without representation. *Artificial Intelligence*, 47:139–159, 1991.

[3] H. De Jaegher and E. Di Paolo. Participatory sense-making. an enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, 6(4):485–507, 2007.

[4] E. Di Paolo. Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, 4:429–452, 2005.

[5] E. Di Paolo and H. Iizuka. How (not) to model autonomous behavior. *BioSystems*, 91(2):409–423, 2008.

[6] D. Floreano, S. Mitri, S. Magnenat, and L. Keller. Evolutionary conditions for the emergence of communication in robots. *Current Biology*, 17(6):514–519, 2007.

[7] D. Floreano and J. Urzelai. Evolutionary robots with on-line self-organization and behavioral fitness. *Neural Networks*, 13:431–443, 2000.

[8] T. Froese and T. Ziemke. Enactive artificial intelligence. *This is a draft paper submitted for journal publication (August 2007). Available on http://www.eucognition.org/white_papers.htm*, 2008.

[9] I. Harvey, E. Di Paolo, R. Wood, M. Quinn, and E. Tuci. Evolutionary robotics: A new scientific tool for studying cognition. In *Artificial Life*, volume 11, pages 79–98, 2005.

[10] E. Husserl. *Expérience et jugement*. PUF (1991), 1938.

[11] K. Manac'h and P. De Loor. A study on the origin of anticipation by guidance for artificial dynamic cognition. In *fourth workshop on: Anticipatory Behaviour in Adaptive Learning Systems (ABiALS 2008).*, 2008.

[12] K. McGee. Enactive cognitive science. part 1: Background and research themes. *Constructivist Foundations*, 1(1):19–34, 2005.

[13] B. McMullin. Thirty years of computational autopoiesis: A review. *Artificial Life*, 10:277–295, 2004.

[14] A. Noë. *Action in Perception*. Cambridge, MA: MIT Press, 2004.

[15] J. Piaget. *The Requilibration of Cognitive Structures.* The University of Chicago Press. The original French edition: L'équilibration des structures cognitives: Problème central du developpement, reprint 1985, 1975.

[16] M. Rohde and J. Stewart. Ascriptional and 'genuine' autonomy. *BioSystems Special issue on Modeling Autonomy*, 91(2):424–433, 2008.

[17] M. Sanchez and M. Slater. From presence to consciousness through virtual reality. *Nature*, 6:332–339, 2005.

[18] J. Stewart and M. Mossio. Is "life" computable? on the simulation of closure under efficient causation. In *ENACTIVE'07 : 4th International Conference on Enactive Interfaces*, pages 271–276, 2007.

[19] Y. Sugita and J. Tani. A connectionist model which unifies the behavioral and the linguistic processes: Result from robot learning experiments. Technical report, Sony CSL Tech. Report: SCSL-TR-00-001, 2000.

[20] M. Suzuki and D. Floreano. Enactive robot vision. *Adaptive Behavior*, 16:122–128, 2008.

[21] J. Tani, R. Nishimoto, J. Namikawa, and M. Ito. Codevelopmental learning between human and humanoid robot using a dynamic neural network model. *IEEE Transaction on systems, man and cybernetics-Part B: Cybernectics*, 38:in press, 2008.

[22] A. Turing. Computing machinery and intelligence. *Minds*, 59:433–460, 1950.

[23] F. Varela. *The Principles of Biological Autonomy*. New York: North Holland, 1979.

[24] F. Varela, H. Maturana, and H. Uribe. Autopoiesis: the organization of living systems, its characterization, and a model. *Biosystems*, 5:187–196, 1974.

[25] L. S. Vygotsky. *Thought and Language - Rev'd Edition*. The MIT Press, August 1986.

Avatar Control and World Interaction in a Non-Visual 3D Gaming Environment

Katarzyna Szymkiewicz * Charlotte Magnusson* Kirsten Rassmus-Gröhn* Emina Karic *
(*)Lund University, Lund, Sweden
E-mail: charlotte.magnusson@certec.lth.se, kasia.sz@gmail.com, kirre@certec.lth.se, emina.karic@gmail.com

Abstract

The present paper reports on the implementation and test of an audio haptic gaming environment. This environment was designed with an ears-in-hand interaction technique allowing turning of the ears in combination with an avatar separate from the haptic interaction point. The purpose of this design was to investigate how the additional rotational degrees of freedom influenced the interaction. The results of the study showed that the additional rotation was sometimes perceived as confusing, although we note that an ears-in-hand tool that can be toggled on or off could still be useful. We also report other qualitative observations which may have influence on the future design of this type of environments.

1. Introduction

For the blind and visually impaired the development of computer graphics has made much information less accessible as most information is presented graphically. The advanced 3D games of today are despite advanced audio capabilities not very accessible due to the manner in which sound is used. Sound effects may do well for enhancing the game experience for a sighted player but to support navigation for a visually impaired gamer the sonification of the virtual world needs to be either realistic or designed with care. Affordable haptic devices have the potential to make, not only the virtual realities of computer games, but any complex graphical environment more accessible than ever. The challenges of future research is how to best map visual representations to audio and haptic feedback and how to make these modalities self-sufficient so that impaired users can begin to enjoy the mainstream games available to their able-bodied peers. Blind accessible games like Terraformers, AudioQuake or the audio game Sarah and the castle of witchcraft and wizardry make excellent use of audio cues to support navigation. Audio - haptic games are still less mainstream. Studies of audio haptic game environments include a Pacman game developed to investigate combinations of sound and haptics in navigational tasks [1]. The avatar (Pacman representation) was positioned at the tip of the the stylus of the PHANTOM device and guided around a maze that was within the PHANTOM workspace. As the avatar was the interaction point, attaching the listener to the stylus position (ears-in-hand) proved useful when the player monitored the position of the ghosts using 3D-sound. The audio haptic game environment developed within the European project GRAB [2] provided a different player perspective. The game was an indoor search and adventure game where the game world also was within the workspace of the interaction device which was a two-finger haptic interface. The user located elements such as bombs, a trap of attracting force, deactivator buttons and a door key in the two-room environment by using two interaction points. Zooming, panning and constrained movement were excluded from the game design. Several panning techniques were developed for the game Haptisk skattjakt (Haptic treasure hunt) [3] where a game world larger than the workspace was explored with the PHANTOM stylus. More extensive tests on navigation and interaction in haptic environments have been performed in virtual traffic environments [4] [5], where techniques such as pushing a limiting box, pressing keyboard arrow keys to move the world in the direction of the arrow or using the button of the PHANTOM stylus to click and drag the world were investigated. All of the scroll functions proved useful and the findings of the tests suggested that different ways of scrolling and zooming can be preferred by different users when performing different tasks. A different set of controls for a game environment using the first person perspective includes the work of Johnston [6] [7]. PHANTOM movements forward and backwards were used to walk the avatar and left and right PHANTOM movements were used to rotate the avatar. A hemispherical dead zone of control in the centre allowed the avatar to stand still. When pressing the button of the Omni PHANTOM device the player could explore the surrounding space in a free explore mode. The work also included navigational aids for the visually impaired composed of a sinusoidal wave force to the stylus and a tapping sound. A zero amplitude to the force and zero volume for the sound was used to indicate the correct direction of the target object. The evaluation suggested the audio aid was most useful.

Other studies of navigational tools utilizing 3D sound and haptics have been made [8] [9] that included attractive forces, a linear fixture, a search tool of crossing planes combined with sound feedback from one object or all objects simultaneously in different combinations. The ears-in-hand metaphor was found to be a useful interaction technique in these environments where the avatar and the listener was positioned at the haptic interaction point (PHANToM proxy) and a fixed forward-facing orientation was used. As yet there have been no studies of the ears-in-hand interaction technique allowing turning of the ears in combination with an avatar separate from the haptic interaction point. This is why we decided to implement and test a gaming environment with this type of design.

2. Implementation 1, the introduction room

The game prototype implements haptic one point interaction from a first-person perspective where the control of the user representation in the virtual world, the avatar, and world interaction is separated. To really test exploration and navigation, the game world was designed to exceed the reach of the PHANToM work space. To familiarize users with the system, the initial environment is a small unfurnished environment as shown in figure 1. The avatar body is represented by the pink vertical cylinder in the middle of the room reaching from the floor to the ceiling.

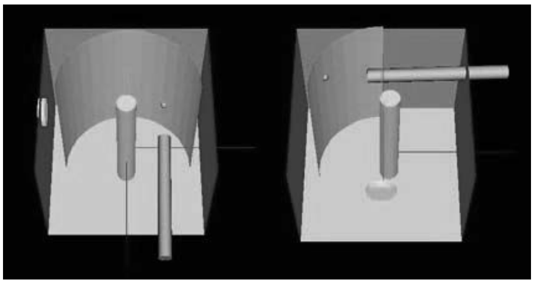


Figure 1: The introduction room from observing viewpoints (the front wall, ceiling, and occlusive parts of the stylus reach boundary are rendered transparent).

The stylus interaction point (shown as a small grey ball in figure 1) was designed to be relative to the avatar body, in an attempt to simulate a blind person's cane (the grey cyclinder in figure 1). To make it easier to separate real objects from the physical constraints limiting the PHANToM workspace, the area where the PHANToM stylus can be moved is constrained virtually. This constraint is shown by the large red half-cylinder in figure 1. Movement of the avatar body, boundary and stylus representation, was controlled by the arrow keys which frees the stylus for use in game world interaction.

The up and down keys walk the avatar forward and backward; the left and right arrow keys turn the avatar left and right. Movement and rotation of the avatar occurs in small steps, but holding down the keys will cause continuous movement. When the avatar approaches a wall (shown in blue in the figure above) it will come within range of the stylus, allowing the user to feel the approaching wall step by step. Once the avatar body is close enough to the wall, the avatar will stop or, if the wall is approached at an angle, slide forward along the wall while emitting an "ouch" sound.

When the stylus is in contact with a wall, turning towards the wall is not possible. Instead, the user is alerted of the obstacle by a "beep" sound. There are two different beep-sounds that are played depending on if the colliding wall is on the left or right of the stylus.

In addition, the room contains a single sound emitting object in the form of a ticking clock (shown by a green disk in figure 2). The listening point is placed at the tip of the stylus, enabling the player to determine the location of the sound source by moving and turning the stylus. The clock has a sound range defined by an ellipsoid, as shown by the black line in figure 2.

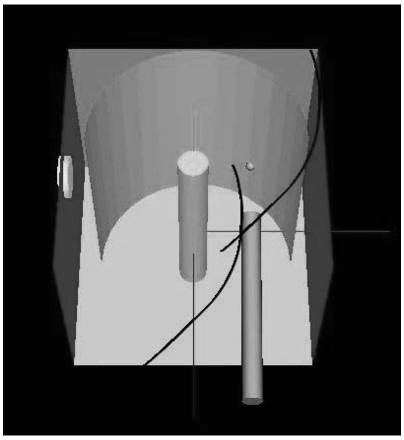


Figure 2: Sound range in introduction room.

It should be noted that it is not only how close you are to the sound that can be detected with the stylus. The entire soundscape is controlled by rotation of the stylus, meaning that if you point the stylus directly at a sound source, the sound will originate from the center speaker; turning the stylus 90 degrees left will cause the sound to come from the right speaker when using a surround system; and so on. This is an extension of the "ears in hand" metaphor shown to be effective in previous applications, since the soundscape now also can be rotated by the PHANToM movements (previously the listener was always assumed to be facing forwards in the virtual environment.

3. Implementation 2: the game environment

This environment consists of a square room furnished with a sideboard, a wall mounted shelf and a table placed alongside the walls.

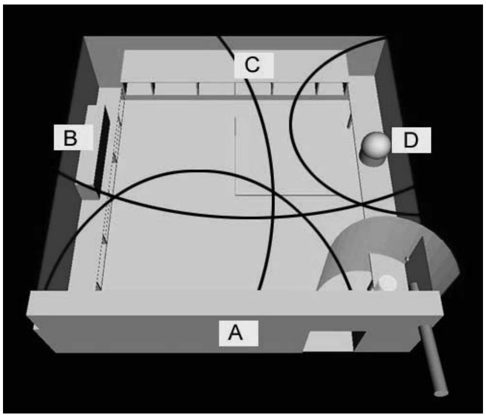


Figure 3: The game environment

Fig 3 location	Object	Description of the sound
A	Wall Clock on the south wall	The sound of the ticking wall clock is the same as in the introduction room
B	Aquarium standing on the side board up against the west wall	The sound of the aquarium is actually a sound capture from a bio lab. It features a bubbling sound for about 6 seconds and continues with the sound of running water.
C	Radio in one of the wall shelf compartments on the north wall	The radio sound is represented by Mozart's *Eine kleine Nachtmusik*
D	Table Clock standing on the table up against the east wall	Compared to the wall clock the table clock ticks more slowly and has a lower pitch.

Table 1. Sound sources

There are four sound emitting objects in the room (see table 1). In the south wall there is also a door frame. The room and the approximate range of the sound sources can be seen in figure 3.

4. User tests

Four visually impaired participants (aged 18, 18, 20 and 27), with previous experience with the PHANToM device, took part in the tests. The test environment can be seen in figure 4.

Figure 4. Test setup.

The test started with a session in the introduction room followed by a test in the full environment. The aim of the first session was to familiarize participants with the controls, the haptics, the sounds of the collision detection and ears in hand by performing a set of tasks:

- Task 1. The test person was asked to walk straight ahead. This resulted in the avatar hitting the front wall, causing the "Ouch" sound to be played. The next instruction was to back up as much as possible, resulting in another "Ouch" sound as the avatar hit the back wall.
- Task 2. At the start of the task, the avatar was positioned with its back against the south wall. The sound of the clock was turned on and the participant was asked to locate the clock. If the participant experienced the beep sounds that indicated restriction of rotational movement while the stylus was in contact with a wall, this was explained by the test leader.
- Task 3. When standing in front of the clock the participant was asked to turn the back to the clock, to turn so that the clock was on the left hand side, to turn so that it was on the

right hand side and then to face it directly. Lastly the participant was asked to make a 360 degree turn i.e. to face the clock again.

- Task 4. The avatar was positioned in the middle of the room, facing the clock. The participant was then asked to locate the four corners of the room, one by one.

After this the participants tried the game environment. The avatar was initially positioned in the southeast corner of the room (by the door). The participant was informed of now being in a different room containing four sound emitting objects and was asked to find them and point them out with the stylus.

5. Test results

In general the users were able to complete the tests quite well. The results of the tests are summarized in table 2 showing task completion (Yes/No) and time for completion in minutes. All users completed the first task of walking back and forth. All users were also able to successfully locate the clock (task 2) although one of the users needed the hint "it may not be on the floor" before finding it. The tasks involving rotating (task 3) and finding corners (task 4) caused some problems.

Task	User 1	User 2	User 3	User 4
1	Y:-	Y:-	Y:-	Y:-
2	Y:0.52	Y:4.20	Y*:2.32	Y:1.10
3.1	Y:0.25	Y*:2.31	Y:1.21	-
3.2	Y*:0.30	Y:0.27	Y:0.22	N:0.41
3.3	Y:0.27	Y:1.00	Y:0.27	N:1.46
3.4	Y:0.31	Y:2.35	Y:0.35	-
3.5	-	Y:0.57	Y:0.49	Y:1.45
4.1	Y:0.56	Y:1.03	Y:0.34	Y:4.20
4.2	Y:0.43	Y:1.26	Y:0.08	Y:0.44
4.3	Y:0.26	Y:1.24	N:0.15	N:1.20
4.4	Y:0.16	-	N:0.01	Y:1.50
Total time	17	32	27	31

Table 2. Task completion summary. Asterisk (*) shows that the task was completed during a second attempt.

Some observations, based on interviews with the participants:

⇒ In general the users tended to keep the PHANToM quite still while walking. The general strategy appeared to be to concentrate on one type of interaction at a time.

⇒ It was interesting to note that the avatar was often referred to as "he" but at the same time the users tended to say "I" when they were touching things with the PHANToM. Also the avatar generated emotional type responses – two users thought he was complaining too much when he hit the wall, whilst two felt sorry for him.

⇒ The fact that the PHANToM stylus could rotate the soundscape complicated matters on several occasions – if you were standing with the clock to the left and then turned the stylus towards it the sound would be heard in the middle.

⇒ It is a problem you can't touch the walls while rotating (this is to prevent falling through the walls).

⇒ It is hard to judge your speed of movement – although the sounds may help.

In the game environment, all four users found the wall clock, the table clock and the aquarium. Only user 4 found the radio. The times used for this test were 22, 26, 34 and 26 minutes. Some observations, again based on interviews with the participants:

⇒ In general the users were able to quite easily find the clocks and the aquarium.

⇒ Also here the users tended to keep the PHANToM quite still while walking. The general strategy appeared to be to concentrate on one type of interaction at a time.

⇒ The fact that you cannot walk sideways makes it harder to scan the room systematically

The radio was very hard to find due to the fact that the sound source did not vary much enough as you moved the stylus. The music was dynamical, and the changes that could have been induced by the motion drowned in the variations in the music itself.

6. Discussion

The current study tested the "ears in hand" for the case where the rotation of the stylus also rotated the soundscape, and where the user had an avatar separated from the PHANToM proxy. In our previous studies [5, 8, 9] we had always used the PHANToM proxy as the avatar, with a fixed facing forwards listener orientation. We wanted to investigate if the introduction of a rotational degree of freedom could be useful. Observations during the test made quite clear that this was not generally the case – instead the rotational degree of freedom introduced problems. One problem was that users could easily rotate the stylus slightly without noticing it. Another problem was that not all users realized the influence of rotation. With stylus rotation of the soundscape there are really two orientations to keep track of – the orientation of the avatar and the orientation of the stylus. Both during this test (and the pilot tests that preceeded it) it became clear that it is probably better to adopt the previously used fixed facing forwards rotation of the listener for the ears in hand design.

Another issue is the design of the workspace (the "reach" of the avatar). It is clear that the surfaces used were too similar and users confused the workspace limits with objects. Also the fact that the workspace contained corners added to this problem. This can probably be remedied by adding textures either to objects or to the limitation of the workspace or possibly by changing the workspace limitation surface to more of an elastic membrane type. And of course by using a more ellipsoid shape to avoid sharp corners.

At the same time the limited workspace can help to show sighted persons some of the problems facing blind or visually impaired persons in their everyday life. One further issue that is to some extent illustrated by the above test results is the fact that a person may perform quite well in an environment also if that person never builds up a cognitive map of the actual layout of the environment. This may be said to relate to navigation by waypoints/routes versus map navigation – and it is well known that it is quite possible to navigate quite well also without an overview mental map of the environment. Still, the test results on this point should be used with caution – just because a person does not remember the layout of the environment after the actual test does not necessarily imply that he or she did not have some notion of it while being in the environment (we saw the same effect in [4] where persons often navigated the environment well, but when they were asked to draw a map afterwards they found this very hard to do).

If we look at the way the test persons used the stylus the general trend was that for walking they would keep it fairly still in front of the avatar while moving the avatar using the keyboard. Also during avatar rotation the PHANToM stylus was in general held quite still. So, in most cases our users preferred to focus on one type of interaction at a time, i.e either they would scan using the PHANToM or they would move the avatar using the keyboard. It is possible that this is something that may change with experience – after all it takes some time to train persons with visual impairments in the use of the white cane.

For the avatar navigation, it was suggested that one should also be able to walk sideways. While it is known that this may make navigation easier (e.g [12]), the current design was for a gaming application where this type of difficulty may actually be part of the challenge of the game. So far we have decided to keep the number of controls to a minimum, but to add sideways movements is clearly an option.

The presented environment illustrates the importance of active exploration. It also illustrates the importance of the close link between gestures and feedback. This is shown both by the difficulty of judging the speed of the avatar movements and the difficulty of finding the radio.

The difficulty of judging the speed of the avatar makes it clear that more feedback is needed for these actions. Auditory cues such as footsteps are suggested, and one could also add some kind of ribbed or checkered texture to allow the user to get movement information through the haptic sense.

The difficulty in finding the radio illustrates what happens when variations in the feedback received is not linked closely enough to variations in the gestures. Since the music played by the radio was quite dynamic, it was hard to really identify the volume changes caused by movement of the stylus. Also, the sound of the radio was heard over quite a large area, and the amount of volume change caused by moving the stylus was not large enough. Furthermore the music piece used was less well suited for spatial hearing – it had the kind of sound that can be a bit hard to locate precisely in space (in the previous navigational study [8] we used a not very dynamic rock'n'roll piece with a rather "heavy" fuzzy guitar since this type of music was seen in pilot tests to make it easy to hear the spatial location).

7. Conclusion

This test shows that this type of environment where avatar control and world interaction is separated can be used also by blind persons. That blind persons are able to handle this type of separated design has also been shown in a recent study where the goal was to help blind sailors to get access to nautical maps [11], which lends further support to our claim.

The rotational degree of freedom introduced made it very difficult to envision the surroundings if the player was identifying himself with the avatar since the soundscape turned with the stylus. If the avatar was facing a sound source and the user pointed the stylus forward, the sound would emanate from the center and front speakers. If the stylus was then turned 90 degrees left, the sound would emanate from the right speakers even though the sound source was in front of the avatar. There can be situations when the direction of the sound is confusing even if the stylus is held pointing forward (facing the same direction as the avatar). For instance, when the sound source is located between the stylus and the avatar on a straight line in front of the avatar the sound will emanate from the rear speakers although the sound source is in front of the avatar. In the prototype the problem was not prominent since all sound emitting objects were placed along the walls.

However, the ears-in hand interaction technique was also seen to be useful when the player focuses on locating a sound source within reach while ignoring the avatar body (although the rotational degree of freedom still is a problem). The ears-in-hand sound would then guide the player towards the target. This leads us to believe the ears-in-hand interaction technique, used as a tool that can be toggled on and off at will, may be valuable also for this type of design where the avatar is separated from the haptic interaction point/proxy. It provides similar properties as attractive forces pulling the stylus towards objects of interest; yet it is superior in the aspect of being able to convey the directions of several sound sources at the same time.

8. Acknowledgements

The study was carried out with the financial assistance of the European Commission which co-funded the NoE ENACTIVE - “ENACTIVE Interfaces”, FP6-2002-IST-1, No IST-2002-002114. The authors are grateful to the EU for the support given in carrying out these activities.

References

[1] Preliminary tests of a Pac Man game, 2007, MICOLE Deliverable D5:Report on the design and evolution of haptic focus and context displays, p.13-15. Available form: http://micole.cs.uta.fi/deliverables_public/deliverables/MICOLE-D5-final.pdf (Accessed: 12 October 2007)

[2] J. Wood, M. Magennis, E. F. C. Arias, T. Gutierrez, H. Graupp, M. Bergamasco, The Design and Evaluation of a Computer Game for the Blind in the GRAB Haptic Audio Virtual Environment, Proceedings of Eurohpatics 2003

[3] Anna Lindholm, 2003, Haptisk skattjakt, Master's Thesis, Certec, Department of Design Sciences, Lund Univeristy, Sweden, p20-21. Available from: http://www.certec.lth.se/dok/haptiskskattjakt/haptiskskattjakt.pdf (Accessed 26 June 2007)

[4] C. Magnusson, K. Rassmus-Gröhn, Non-visual Zoom and Scrolling Operations in a Virtual Haptic Environment, EuroHaptics 2003, Dublin, Ireland 2003

[5] C. Magnusson, T. Gutierrez, K. Rassmus-Gröhn, Test of pan and zoom tools in visual and non-visual audio haptic environments, ENACTIVE 07, Grenoble, France, November 19-22, 2007

[6] Multimodal interaction and control in a virtual environment. J. Johnston, A. Crossan. and S.A. Brewster, Multimodal Interaction and Control in a Virtual Environment, In Vol II Proceedings of the First International Workshop in Haptic and Audio Interaction Design (HAID2006). University of Glasgow, UK, 31st August - 1st September 2006

[7] J.Johnston, 2006, Multimodal Interaction and Control in a Virtual Environment Individual Project Report. Available from: http://www.dcs.gla.ac.uk/%7Ejohnstoj/project/report.pdf (Accessed: 19 October 2007)

[8] C. Magnusson, K. Rassmus-Gröhn, Audio haptic tools for navigation in non visual environments, ENACTIVE 2005, the 2nd International Conference on Enactive Interfaces, Genoa- Italy on November 17-18, 2005

[9] C. Magnusson, H. Danielsson, K. Rassmus-Gröhn, Non Visual Haptic Audio Tools for Virtual Environments, Workshop on Haptic and Audio Interaction Design, University of Glasgow, 31st August - 1st September 2006.

[10] Rassmus-Grön, K., Enabling Audio Haptics, Licentiate Thesis, Certec, Department of Design Sciences, Lund University, Sweden, p33-34

[11] Simonnet, M., Vieilledent, S., Guinard, J-Y., Tisseau, J., Can haptic maps contribute to spatial knowledge of blind sailors?, ENACTIVE 07, Montpellier, 2007

[12] Wallergård, M., Lindén, A., Davies, R. C., Boscian, K., Minör, U., Sonesson,B., Johansson, G., Initial usability testing of navigation and interaction methods in virtual environments: developing usable interfaces for brain injury rehabilitation, Presence: Teleoperators & Virtual Environments . 16 (1), 16-44, 2007

Assistive Technologies as Part of a Digital Ecosystem

David J. Calder
Digital Ecosystems and Business Intelligence Institute (DEBII), Curtin University of Technology, Australia
E-mail: david.calder@cbs.curtin.edu.au

Abstract

There is a rapidly growing requirement for a new approach to the design of computer-based assistive systems for the disabled. Within a disparate classification of user requirements, many systems manufacturers have catered for particular disabilities in relative isolation and not necessarily within the wider context of other modalities and areas of disability. The Digital Ecosystems for Assistive Technologies (DESAT) approach would assist therapists, educationalists, manufacturers and academic researchers, by offering a coordinated digital ecosystem from which both the ecologically embedded end user and support teams from diverse disciplines could benefit.

1. Introduction

Along with a plethora of computer-based solutions to the challenges of a varied and complex range of human disabilities, comes the danger of regarding the client as an isolated and static entity. Many people have one or more progressive disabilities. These clients are dynamically interactive within an ecology that is itself constantly adapting and changing in real time. Far from being static, complex human systems are constantly interacting with one another as well as other machine systems within the ecology. As these human systems evolve, so the whole surrounding ecology evolves too.

A technology-centred approach that produces an inherently inflexible solution will be flawed. Many designs based on requirement specifications for disabled users fall short as ongoing adaptive solutions. In practice, it is currently almost impossible to offer an alternative or augmentative technology that will match and then continue to match the user requirements of an ever evolving system, as is the case where human input and output channels and the complexities of human-to-human and human-to-machine communication are concerned. The rejection rate of currently produced assistive devices reinforces this claim and is discussed later in the paper. Each individual is a moving target for the system designer, as the day-to-day progression rate for each user is different. The long-term user requirements are therefore dynamic in essence.

There are also factors such as the time and financial investment involved in designing and manufacturing for one person in a bespoke tailor-made fashion. The answer is often a compromise.

These challenges are compounded if the user has multiple disabilities [1]. The design team, who may be focussing exclusively on one area of disability, does not always consider these multi-facetted demands.

Multiple disabilities can be physical, cognitive or both. Issues of complexity with respect to individual requirements must be seen within the context of a wider ecology of the particular user, with that person clearly at the centre, contributing to a team solution. An established and highly successful ecological approach to designing individualized education programmes for the disabled student has been refined over twenty years into a highly recommended model and is now regarded as 'best practice' [2]. This ecological approach has not as yet permeated all areas of disability support. However, the power of the digital ecosystem framework is now accepted within many other disciplines, particularly with respect to small enterprise collaboration [3].

Within small business, the advent of the web has allowed sales penetration over vast distances. Accompanying these advances have come new modes of marketing and partnership possibilities that would have been impossible only a few years ago. With this connectivity has come a fertile and dynamic business theatre that cannot be avoided if small enterprises are to survive. This interaction has led to collaborative workflow models [4].

The logic behind collaborative workflows is to produce a sequence of activities that not only produce a meaningful result, but also to facilitate small groups working together to achieve common goals. The actual physical distance and associated limitations between these entities then becomes less important as web based tools are used to link enterprises and their common aspirations [5]. The entities themselves may be small companies competing against large predator corporations, or widely dispersed cottage industries (such as those associated with assistive devices) with a common interest [6].

Beyond the standard empowerment the digital ecosystem model has provided, are more specific areas that are pertinent to such groups operating in harmony. One of the most important of these is trust evaluation [7]. Other typical support areas are logistics and privacy [8, 9]. These would act as foundations for the model that is proposed in this paper.

Digital Ecosystems for Assistive Technologies (DESAT) is a collaborative cluster-based ecosystem, neither limited by distance between clusters nor the particular disability types associated with each of the clusters. Individual clusters may include a range of specialist personnel associated with the support of a client requirement. The output of such an environment would not only be the efficient research and development of appropriate assistive devices, but also result in more streamlining for the teams in their everyday support of an individual, be that speech therapy for dysarthria patients or training in the use of a long cane or mobility aid for the visually impaired.

2. A Range of Disabilities

With respect to vision perception in particular, there are a number of theories that complicate the challenges with respect to assistive systems design [10]. The machine vision computational theories of David Marr [11] do not equate with the constructive theories of Helmholtz and Rock [12]. None of this uncertainty makes matters easier for assistive device research and design personnel.

When any one of the major human system input channels is affected, the functional potential of the whole person will be compromised. A loss of vision in particular, presents extraordinary challenges. With time, other viable modalities have been shown to compensate and become enhanced in sensitivity. This enhancement occurs in the compensating areas of the brain [13]. With time, the human system adapts and compensates for the loss of a modality, to some extent at least. For the assistive device engineer, however, there is a continually moving target – the requirement specification is therefore constantly changing and never fixed. Recent advances in the understanding of brain plasticity reinforce this perspective. It has been demonstrated that driving the brain with demanding sensory, cognitive and motor activities on a frequent basis will often result in a positive outcome. The reverse is also true [14].

Whereas vision impairment may often be linked to the input channel that results in sight, physical disability may also be associated with the output channels from the human system. The loss of motor speech is often associated with paraplegia or quadriplegia. Communication and mobility may be simultaneously compromised.

It is generally accepted by interventionists that an early programme must provide the learner of communication skills with a systematic means of restoring and maintaining control over his or her environment. Augmentative and alternative communication specialists have for many years, made use of an ecological model to promote control for the disabled learner of communication skills [15].

The above refer almost exclusively to a means of acceptable two-way communication for the individual within a wider environment. The same individual may also require human and technological support to move around the home and to control that encapsulated environment, let alone the wider environment.

3. Cognitive disability

Physical disability is usually, though not always, accompanied by varying degrees of cognitive impairment. For support teams, whether they are educational psychologists, occupational or speech therapists, computer hardware or software engineers researching and developing assistive devices, the task is formidable.

An early introduction to one of the challenges was faced by the author in designing a portable computer-based speech aid [16], [17]. As long as the subject using the working prototype actually knew what he or she wanted to say, the system would allow a range of alphanumeric or iconic software generated key input options to be used or even a single switch input activated by an eyelash. However, as cognitive impairment often accompanies speech loss, even simple iconic prompts can prove to be futile in the interface. Even today, this problem still rules out the use of a speech aid for many people who have no speech. Although a touch screen interface is usually standard, an essential option is the remote single and dual switches. These cater for a person with quadriplegia who may have limited movement, such as the muscles of the eyelids. Stored interface display options include both written cues and a variety of rewritable iconic keys, allowing for some tailoring to the cognitive state of the patient or user. Nevertheless, the user has to have good cognition and particularly an intact short-term memory as a prerequisite!

4. A Spectrum of Assistive Devices

Categories of assistive technologies generally fall into groups that cater for the impairment or loss of a single modality. Cochlear implant and hearing aid researchers and manufacturers tend to cluster around the associated network of audition therapists, educationalists and even competitors. Assistive devices for the blind or partially sighted fall into a different camp. The more advanced mobility aids for the blind and implant units for the deaf use digital technology, both requiring miniaturisation, precision and robustness. However, the teams working within their respective boundaries, may not be sharing information that could be to the good of all and particularly those disabled people who will be using these technologies.

The author's research has been interdisciplinary but primarily based within engineering and applied science. It has been noted that a software application such as our pronunciation tool for use within speech therapy [18] and speech aid devices (already cited) encompassing both hardware and software design, have some overlap in clusters of interested parties.

However, our current prototype mobility aid for the blind, relates to a completely distinct and different cluster, none appearing to optimise resources and communication with the others. For example, the speech therapists and support teams for a cardiovascular accident victim, who may be likely to use speech therapy software or introduce a portable speech aid, are not readily associated with those comprising a support team for the visually impaired client. Certainly the design and manufacturing teams for these aid examples have been, for the most part, historically separated.

A mobility aid for the visually impaired is a portable electronic device that is either hand-held or worn by the client, which warns of obstacles ahead. These devices suffer from a number of problems, the most important of which are related to the interface that conveys information to the user. Most aids use vibrating buttons or sound alerts to warn of upcoming obstacles, a method which is only capable of conveying very crude information regarding direction and proximity to the nearest object. Some of the more sophisticated devices use a complex audio interface in order to deliver more detailed information, but this often compromises the user's hearing, a critical impairment for a blind user.

The World_Health Organization estimated that in 2002 there were 161 million (about 2.6% of the world population) visually impaired people in the world, of whom 124 million (about 2%) had low vision and 37 million (about 0.6%) were blind [19].

Projections indicate that by 2024, over 800,000 Australians will suffer from visual impairment, and approximately 90,000 will be blind [19]. Vision impairment is responsible for 18 percent of hip fractures by older Americans at a cost of treatment of $2.2 billion each year, according to the Framingham Eye Study. If we could prevent just 20 percent of such hip fractures, it is estimated that US$441 million would be saved annually [20].

5. Rigid Designs and Failure of Acceptance

It is now accepted by Human Computer Interaction (HCI) specialists, that earlier user interface designs were often driven primarily by technology rather than being user centred. Although challenging interface designs [21] require the guidelines and standards of a user centred methodology [22], the results with respect to their implementation and user acceptance are not that encouraging.

An extreme example of a technology-driven user interface design is the commonly used QWERTY keyboard layout [23]. Designed to slow input from the user in order to prevent the (mechanical) typewriter hammers from jamming during operation, the result was constrained by engineering limitations of the time.

Sophisticated and very expensive *cutting edge* electronic travel aids may use sound patterns as a substitution for sight. As proof of a significant user interface problem with some of these, specific examples can be found from Johnson and Higgins who refer to visual–auditory substitution taxing a sensory modality that is already extensively used for communication and localization [24]. Velazquez [25] refers to four shortcomings of existing ETAs. One of these is they provide an acoustic feedback that *interferes* with the blind person's ability to pick up environmental cues through hearing.

Recent studies indicate that a 20 minute usage of acoustic feedback devices causes serious human information registration, reduces the capacity to perform usual tasks and affects the individual posture and equilibrium [26]. Many audio sensory substitution devices fail because of their complex, confusing and restrictive audio feedback to the user, which blocks natural ambient sounds. They are therefore not suitable for a typical blind user who will probably have multiple disabilities. A Study by Ross and Blasch [27] clearly indicated that blind people preferred a simple tapping tactile interface to a device generated sound feedback!

The digital ecosystem paradigm offers opportunities for both knowledge sharing within a wider ecology as well as user cognition-centred adaptability and flexibility for all assistive technologies.

6. Aims of the DESAT Model

With each client representing a nucleus at the centre of his or her support cluster, an individual's local ecological environment has been acknowledged (as discussed and cited in previous sections) as a worthwhile starting point, offering a framework from which specialist support action may be fleshed out.

Each support cluster would have a number of clients within its particular category of disability. Cluster membership would not be determined by distance or physical boundaries. The aim would be to maximize use of the digital ecosystem paradigm in order to break existing physical boundaries.

By applying a DESAT strategy, current digital technologies such as mobile, the internet and video conferencing can be coordinated and optimised to deliver the best outcome for all members of this ecosystem.

Open-ended but novel design solutions would be encouraged from both hardware and software developers. The sharing and exchange of common modular solutions at both a functional and user interface level would be part of the ecosystem membership requirement. The protection of intellectual property (IP) would remain an individual company's prime commercial consideration. The difference would be in the focus and modular consideration of appropriate novel and relevant ideas, when first considering I.P. matters. This will not always be relevant to designs, but when it is, it should in fact enhance the potential for profit and sales within the DESAT community itself, as well as in a wider context (external to the ecosystem).

Those academic cluster members who currently work within a limited research environment with a very small interest group would have the opportunity to share their research and ongoing projects on a wider stage within the digital ecosystem. Cross-disciplinary interaction would be nurtured by DESAT. Members of diverse university schools such as computer science, electronics, psychology and education would be made aware of significant ongoing DESAT research and those projects requiring urgent and immediate input from members.

7. Outline of the DESAT Structure

A cluster of people with a vast range of interdisciplinary skills would focus on a user group of people all with a common disability. There would be many separate clusters, meeting the challenges of specific needs of different disability groups. As now, it may be assumed that special education specialists, therapists, medics, academics, engineers and particularly hardware and software experts would form part of each cluster, the main difference being a recognition of the greater ecosystem in which each cluster coexists and operates.

Users at the center of each cluster, the nucleus, would determine the nature of the environment. Clusters would communicate with each other for a common good and the ecosystem itself would be able to benefit from its size in terms of external links and its critical mass. See Figure 1.

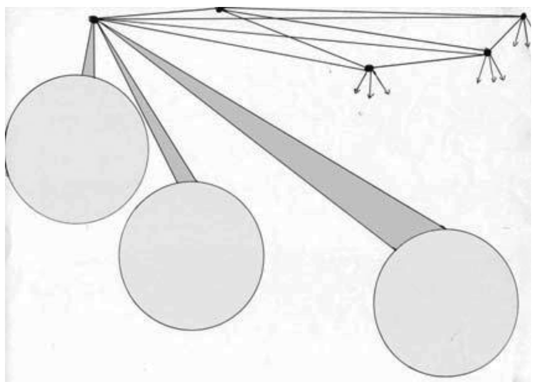


Fig. 1. DESAT structure showing clusters

A starting point for such a structure may take into account the problem as defined by Liu et al when referring to *building the right systems* and the need for better tools and environments in their paper on component-based medical and assistive devices and systems [28]. They put forward a ten-year roadmap, which fits well as a start to implementing the DESAT paradigm. Clusters need to be client centered, taking into account breakthrough research such as that of Bach-Y-Rita into sensory substitution [29] and Merzenich into brain plasticity [30].

8. Operational Benefits

A global advantage and DESAT's greater mass would benefit the ecology on many levels. There would be lower manufacturing costs than is now associated with small-run dedicated systems production. This advantage would result from greater demand for DESAT modular units across clusters and existing boundaries. Relatively large production runs catering for a global DESAT module demand would drive production costs down.

At an academic level, research on large consensus-driven projects could be split, not only between disciplines, but also across campuses and national boundaries. By optimising funding with coordinated grant applications, work need not be duplicated. Projects involving third year undergraduate as well as postgraduate students would have a number of specific benefits. An early introduction of DESAT to students would provide a pathway and assist in nurturing future scientists and other specialists within a variety of disciplines.

DESAT would make use of new and growing technologies such as high speed internet and video conferencing links, mobile phone and other miniaturized digital platforms.

There would, however, be a difference in the way all these technologies are harnessed. Fast and efficient communication channels would be underwritten by each member's awareness and acknowledgement of specific ecosystem guidelines. These guidelines may later form appropriate standards.

9. Guidelines and Standards

Standards institutions such as the International Standards Organisation (ISO) have produced specifications for both hardware and software. As the ecosystem will involve both hardware and software and many existing standards and guidelines already apply, it may be assumed in the early stages at least, that specific DESAT guidelines would also have to be drawn up and adhered to within the boundaries of the fledgling ecosystem.

Guidelines for software and user interface design [31] have been available for many years and offer a good starting point for the conceptualisation of this process. There is no doubt that recent web design standards such as those from the National Institute of Standards and Technology (NIST) work well in that particular context and have indeed been accepted. However, international assistive technology standards are not always adhered to, as they are developed largely for voluntary compliance [32].

In his text on software development ecosystems, Highsmith [33] refers to the dangers of rigid development team processes and the need for a balanced view. Maintaining a team in a generative state where innovation can flourish, requires 'balancing at the edge of chaos'. An excess of order causes stasis whilst an excess of chaotic latitude results in a degenerated random outcome. A compromise balance of the two extremes appears to be needed for the best outcome. Highsmith labels this ideal balanced state as CHAORDIC.

Applying rigid standards and guidelines to DESAT may result in an equivalent ecosystem stasis. A CHAORDIC balance, similar to the above, may prove to be the ideal: particularly in terms of enhanced productivity, considering the dynamic nature of the proposed ecosystem.

10. Conclusion

An ecosystem paradigm has been proposed in order to optimise strategies aimed at improving the quality of life for people with a range if disabilities. The DESAT ecosystem is made up of dynamically changing entities, client-support clusters being the most important. Clusters would be client-centred, with an interactive group representing the nucleus and contributing to a particular cluster direction.

Use is made of digital technology within the ecosystem to help break many conventional restrictive cluster barriers and attain the best outcome. An open-ended and reciprocal modular manufacturing approach has been suggested within the overall boundaries of the ecosystem. Typical production deficiencies in the small-run manufacture of assistive devices may be reduced or overcome. The current concept-to-customer delivery time and cost could be reduced. The dynamic requirement of this model necessitates a fluid approach to guidelines. A balanced structure will be required in order to best harmonise the productivity and overall function of the ecosystem.

11. References

[1] J. Jacko, H. Bautsch, "A Conceptual Framework For The Integration of Clinical Diagnosis, Functional Capabilities and Classes of Technology", *University of Wisconsin Lab for Human-Computer Interaction.* Technical Report series no. 1131, 2000, pp. 1 – 59

[2] B. Rainforth, J. York, C. Macdonald, *Collaborative Teams for Students With Severe Disabilities,* Baltimore: Paul Brookes, 1993, pp. 71- 83.

[3] E. Chang, M. West. "Digital Ecosystems and Comparison to Collaboration Environment". *WSEAS*

Transactions on Environment and development 2, 2006, pp. 1396-1404

[4] L. Pudhota, Chang E, "Modelling the Dynamic Relationships between Workflow Components" ICEISI Porto, Portugal, 2004.

[5] D. Neumann," An Introduction to Web Objects". On-line at http://mactech.com/articles/mactech/Vol.13/13.05/WebObjects Overview. 2004

[6] M. UlieruR. Brennan Scott*, "*The Holonic enterprise: a model for Internet-enabled Global Manufacturing Supply Chain and workflow Management*",* Canada, 2000

[7] E. Chang, T. Dillon and F Hussain, "Trust and Reputation for service-oriented Environments: Technology for building Business intelligence and Consumer Confidence," John Wiley and Sons, West Sussex, England, 2006

[8] M. Clark, P. Fletcher et al. Web Services Business Strategies and Architectures. Expert press. 2002

[9] G. Skinner and E. Chang. "A Projection of the Future Effects of Quantum Computation on Information Privacy and Information Security" *International Journal of Computer Science and Network Security 6,* 2006, pp166-172"

[10] J S. Coren, *Sensation and Perception,* Fort Worth: Harcourt Brace, 1994, pp. 328 –360.

[11] D. Marr, *Vision.* San Francisco: W.H. Freeman, 1982.

[12] I. Rock, *The Logic of Perception,* Cambridge: MIT Press, 1983

[13] E. Goldstein, *Sensation and Perception,* Boston: Brooks Cole, 1999, pp. 377 – 378.

[14] M. Woodlee, T. Schallert. "The interplay between behavior and neurodegeneration in rat models of Parkinson's disease and stroke", *Restor Neurol Neurosi,* 2004, 22(3-5); pp 153 - 161

[15] J. Reichle, J. York, T. Sigafoos, *Implementing Augmentative and Alternative Communication – Strategies for Learners with Severe Disabilities,* Baltimore: Paul Brookes, 1991, pp. 61 – 70.

[16] D.J. Calder, P.F. Lister and J.D. Mck Watson, "Large Vocabulary Synthesized Speech Aids for the Vocally Handicapped," *Augmentative and Alternative Communication,* vol 2, no 3, pp 98, Aug. 1986.

[17] D. Calder, "Flexible Speech Aid Designs Requiring Iconic Soft keys", in *Proceedings of the 5th International Symposium on IC Technology, Systems & Applications, 1993, pp. 451 – 456.*

[18] D.J. Calder, B.M. Chen and G.G. Mann, "A Multimedia Training Tool For Speech Impaired Clients," in proceedings of ED-MEDIA 96 World Conference on Educational Multimedia and Hypermedia, 1996, pp 69 – 75.

[19] World Health Organization: Magnitude and causes of visual impairment, p.3, Nov 2004

[20] Vision 2020: the Right to Sight 1999 - 2005 p. 19

[21] A. Edwards, *Extra-ordinary Human-Computer Interaction: Interfaces for Users with Disabilities,* Cambridge University Press, 1995.

[22] A.Dix, J Findlay, G. Abowd, R. Beale, *Human-Computer Interaction,* New York: Prentice Hall, pp. 160 - 173.

[23] B. Shneiderman, *Designing The User Interface,* Reading MA: Addison Wesley Longman, 1998, pp. 308 – 311.

[24] L Johnson and C Higgins, "A Navigation Aid for the Blind using Tactile-Visual Sensory Substitution," Department of Electrical and Computer engineering program, university of Arizona, Tucson USA. 2006

[25] R.Velazquez, E. Pissaloux and F. Maingreaud, "Walking Using Touch", *Proc. Of 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference,* Shanghai, China, 2005

[26] J. Hakkinen, "Postural Stability and Sickness symptoms After HMD Use", *Proc. Of IEEE International conference on Systems, Man and cybernetics,* Hammamet, Tunisia, 2002

[27] D. Ross and B. Blasch, "Wearable Interfaces for Orientation and Wayfinding," *Proc. Of ASSETS 2000*, Arlington, Virginia, USA. 2000

[28] J. Liu, B. Wang, H. Liao, C. Shih, T. Kuo, A. Pang and C. Huang, "Component-based Medical and Assistive Devices and Systems," in *Proceedings of High Confidence Medical and Systems (HCMDSS) Workshop,* 2005, Philadelphia, PA.

[29] P. Bach-Y-Rita and S. Kercel, "Sensory Substitution and the Human-Machine Interface," *Trends in Cognitive Sciences,* vol. 7, pp. 541-546, 2003

[30] M. Merzenich, W. Jenkins, *Memory Concepts,* Elsevier, 1993, pp 437 - 453.

[31] D.J. Mayhew, *Principles and Guidelines In Software and User Interface Design*, Englewood Hall N.J: Prentice Hall, 1992.

[32] R. Cooper, "Harmonisation of Assistive Technology Standards," in proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society, Vol. 20, No 6, 1998

[33] J. Highsmith, *Agile Software Development Ecosystems,* Boston: Addison Wesley, 2002, pp. 341

Perceptual Evaluation of a Real-time Synthesis Technique for Rolling Sounds

Emma Murphy*, Mathieu Lagrange*, Gary Scavone*, Philippe Depalle*, Catherine Guastavino*
(*) Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT),
McGill University, Canada.
E-mail: emma.murphy@mcgill.ca, mathieu.lagrange@mcgill.ca, gary@music.mcgill.ca, depalle@music.mcgill.ca, catherine.guastavino@ mcgill.ca

Abstract

In this study 6 different versions of a new real-time synthesizer for contact sounds have been evaluated in order to identify the most effective algorithm to create a realistic sound for rolling objects. 18 participants took part in a perceptual evaluation experiment. Results are presented in terms of both statistical analysis of the most effective synthesis algorithm and qualitative user comments. Finally recommendations for future implementations of synthesis techniques and subsequent perceptual evaluations are presented and discussed.

1. Introduction

The role of perceptual validation as part of any sound real-time sound synthesis process is crucial for the development of useful algorithms to create sounds with "natural" or "realistic" attributes. The application of contact sounds to virtual reality systems should assume a level of realism in order to create convincing environments for the user. Furthermore in the creation of real-time sound synthesis techniques for new interfaces for musical expression or for the design of enactive interfaces, the perception of naturalness is crucial to create intuitive interactions. In this study a number of synthesis technique versions have been evaluated to identify the algorithm that generated the most natural or realistic sound for a rolling object.

As part of the field of ecological acoustics William Gaver has explored issues in the analysis and synthesis of physical sounds to create effective algorithms for the synthesis of basic-level events such as impact, scrapping and dripping as well as more complex events such as bouncing, spilling and machinery [1, 2]. Furthermore the real-time synthesis of contact sounds has received much attention in the auditory display community and some convincing results have been achieved [3, 4].

There is a significant body of research concerning the identification [5] and classification [6] of everyday and environmental sounds. More specifically, Warren and Verbrugge [7] have investigated the perceptual attributes of breaking and bouncing events from a temporal perspective. Van den Doel et al. [8] have investigated measurements of the perceptual quality of sound synthesis for contact sounds. Furthermore Stoelinga [9] conducted auditory perception experiments investigating the perceptual understanding and evaluation of the direction, size and speed of rolling objects.

This present study details the evaluation of different versions of a sound synthesis technique, which has recently been proposed in [10]. Firstly the synthesis techniques are described in relation to the sound set stimuli under evaluation in the present study. The experimental design and method are presented followed by a discussion in terms of qualitative and quantitative results and recommendations for future work.

2. Real-time Synthesis Technique

Algorithms that have been proposed for the purpose of real-time rolling sounds synthesis [11, 12, 4] are based on empirical settings of the parameters of the algorithm. The algorithm evaluated here is an attempt to overcome this limitation by estimating the synthesis parameters from actual recordings. The analysis method follows a standard source/filter approach where the filter parameters are estimated using a High-Resolution modal analysis technique [13]. This method is processed over a recording of the plate hit by the rolling object. The estimated filter parameters are then considered to parametrize a deconvolution filter implemented as a structure of cosine cells [14]. The rolling sound is then deconvolved using this filter in order to estimate the source signal. A more detailed description of the analysis scheme can found in [10]. This signal is then modeled as a series of triggers of an amplitude modulated impact signal. Determining whether the filter should encode only the damping and frequency of the modes alone or also include the gains is a design issue that is difficult to gage during the synthesis implementation process. Therefore the two cases (referred to as gain/ no gain) are considered in the experiment presented in this study in order to evaluate whether one alternative is better than the other from a perceptual perspective.

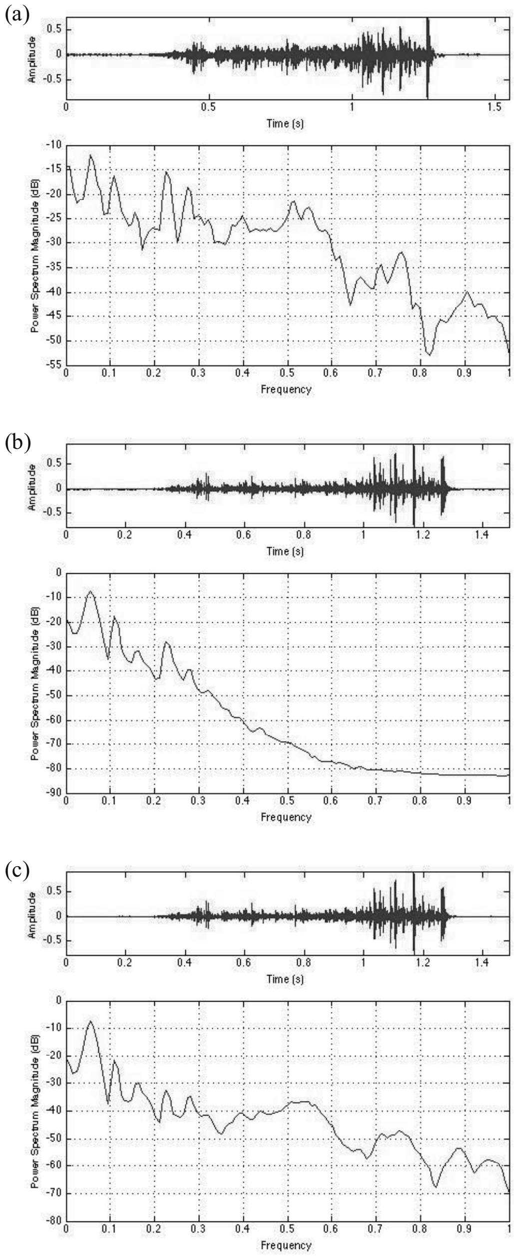


Fig 2: Synthesized sounds of a marble rolling over a highly inclined MDF plate using the three evaluated schemes: (a) impact excitation signal (Raw) (b) Meixner window (c) a combination (Mix)

This impact signal can be of different shapes. One, termed "Raw", is the selected section of the deconvolved signal. Another, termed "Meixner" is the fit of a parametric shape commonly used for the modeling of attacks in the audio coding area, namely the Meixner window [15]. The final signal shape, termed "Mix" is an additive combination of the two previous ones. The intent behind the Mix approach is to balance the properties of the two previous shapes (Raw and Meixner). Indeed, by inspection, it was found that the Raw induces too much high frequency content, whereas the Meixner too much low frequencies, see Figure 1. As shown on the spectral plot of the Mix synthesis, the spectral content is more balanced, approximating more closely the spectral content of the original recording displayed in Figure 2. It was therefore expected that the Mix would be rated rated highest by participants in the perceptual evaluation.

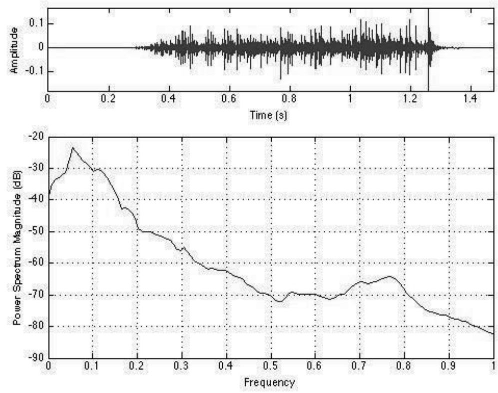


Fig 2: Original recording of a marble rolling over a highly inclined MDF plate

3. Perceptual Validation Experiment

3.1 Stimuli

Impact and rolling sounds were recorded and analyzed to obtain specific model parameters. We considered three different rolling objects: a half liter bottle made of glass (rolling on its side), a small glass marble, and a croquet ball made of wood. The rolling surface was either a medium density fiberboard (referred to as MDF) plate of 95 by 25 by 2 centimeters melaminated or a medium density fiberboard (referred to as Medium) of 80 by 30 by 2 centimeters non-melaminated. Both contact and rolling sounds were recorded, with three different plate inclinations (to vary speed). The sounds were recorded in an IAC double-walled sound isolated booth using both external Behringer omni-directional microphones (ECM 8000).

Table 1: Rolling Stimuli

Synthesis Shape	Filter	Rolling Speed	Material
Raw	Gain	Fast	Bottle, Glass, Wood
	No Gain		
	Gain	Medium	Bottle, Glass, Wood
	No Gain		
	Gain	Slow	Bottle, Glass, Wood
	No Gain		
Mix	Gain	Fast	Bottle, Glass, Wood
	No Gain		
	Gain	Medium	Bottle, Glass, Wood
	No Gain		
	Gain	Slow	Bottle, Glass, Wood
	No Gain		
Meixner	Gain	Fast	Bottle, Glass, Wood
	No Gain		
	Gain	Medium	Bottle, Glass, Wood
	No Gain		
	Gain	Slow	Bottle, Glass, Wood
	No Gain		

Six synthetic versions (3 synthesis shapes with and without gain filters) were evaluated for 2 different plates (MDF and Medium), on 3 different Materials (Bottle, Glass and Wood) at 3 speeds (Fast, Medium and Slow). This gave a total sounds set of 108 sounds; 6 synthesis versions across 18 trials. However one sound set (the bottle rolling on the medium plate at a slow speed) had a significantly lower gain in comparison to the remaining sound sets across the other trials. It was considered that normalizing all sounds could have had an impact on the perceptual judgments of the synthesis techniques under evaluation. Therefore this sound set was excluded from the experiment and users were presented with a total of 102 sounds, six sounds across 17 trials. [1]

The experiment took place in an acoustically treated room and sounds were played back through closed headphones (AKG K271) at a comfortable level on a Mac Pro through a MOTU 828MKII audio interface.

3.2 Experimental Design

18 participants between the ages of 21 and 47 (AV: 27, S.D. 7), students or staff at McGill University were recruited for the perceptual evaluation experiment. On each trial, participants were presented with six sounds. They were asked to first of all listen to all of the sounds and then to rate the extent to which each version sounded like a rolling object. Participants could repeat individual sounds as many times as desired and play counts were recorded. Presentation of sounds within and across trials was randomized.

To indicate a rating for each sound, users were asked to move a slider button for the corresponding sound, over a continuous scale with both numeric (ranging from 0 to 100) and verbal descriptors. Verbal descriptors were adapted from those used in a previous study [16] conducted by van den Doel et al. to evaluate perceptual attributes of liquid sounds. In this present study verbal descriptors for “rollingness” ranged from “Not at all like rolling” (0), “A little bit like rolling” (25), “Somewhat like rolling” (50), “Close to Rolling” (75) to “Exactly like rolling” (100). Participants were encouraged to use the full range of the rating scale. Users were also encouraged to describe the sound or justify their rating in “comments” text boxes provided for each sound (see figure 3).

As many of the synthesized versions for evaluation were not close to the original recordings it was decided that the originals would not be presented for comparison or included as hidden reference signals. Furthermore the aim of the study was to evaluate the sounds in terms of realism rather than replication or quality in comparison to the original. Therefore only the various synthesized versions were presented to participants for evaluation. However participants were also asked to rate the original versions of the sounds as a separate task. This step was taken in case users did not use the full scale to judge sounds then their ratings could then be normalized against their rating of the originals. The testing interface was developed based on a modified version of the MUSHRAM interface [17]). However it should be clarified that the MUSHRA method was not used for perceptual quality evaluation for this experiment. Rather participants were presented with a straightforward comparison task.

In addition to providing verbal comments, participants were asked to complete a post-task questionnaire in which they were asked to describe the differences in the sounds presented in each trial and explain their strategy for rating the sounds.

4. Results

All participants made use of the full range of the rating scales from 0 to 100 and therefore it was not necessary to normalize the ratings using the perceptual data from the original recordings.

The mean results of participant ratings according to synthesis version with standard deviations are

[1] Sounds are available at http://mt.music.mcgill.ca/~mlagrange/enactive/deliverables/2/rolling

illustrated in figure 4. The Meixner shape was rated highest amongst participants and the Raw shape was perceived as lowest on the scale of "rollingness". Interestingly the mean perceptual ratings for the Mix synthesis technique fall between the Meixner and Raw mean ratings. This in fact confirms the actual synthesis process as the Mix shape is a combination of Meixner and Raw shapes.

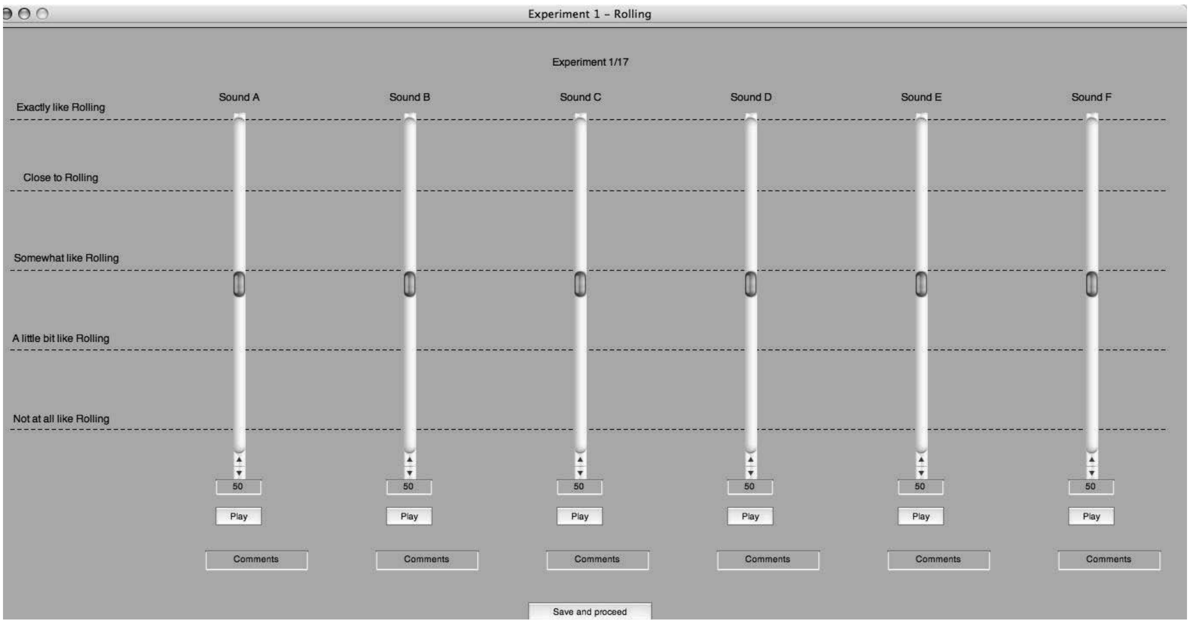


Fig 3: Perceptual Evaluation Testing Interface

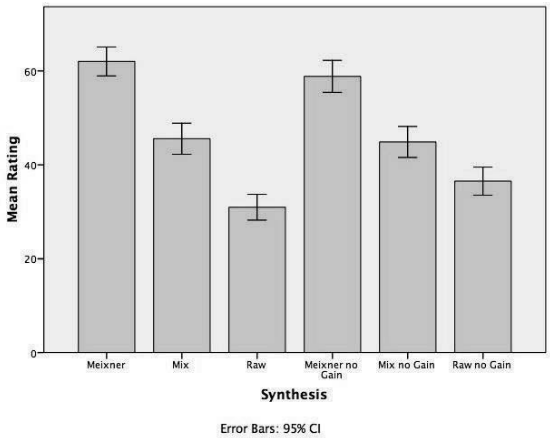


Fig 4: Mean ratings according to synthesis version collapsed over participants, speed and material. A main effect of synthesis shape was observed (p = <0.001) but no significant effect of filter type (gain/no gain).

A 3 (synthesis shape) x 2 (filter – gain, no gain) x 3 (speed) x 3 (material) x 2 (plate) factorial ANOVA revealed main effects of synthesis shape (F(2,1833)=159.4, p = <0.001), material (F(2,1833)=63.4 , p = <0.001), and speed (F(2,1833)=10.7 , p = <0.001) but the effect of both plate and filter (gain/no gain) were non-significant. The following interaction effects were observed among variables: synthesis shape * material (F(4, 3668)=10.5, p = <0.001), synthesis shape * speed (F(4,3668)=4.0, p = <0.01), speed * material (F(4,3668)=6.5, p = <0.001), filter * speed (F(2,3669)=17.5, p = <0.001) and filter * material (F(2,3669)=7.0, p = <0.01). No other interaction effects were observed, specifically no interaction of plate with any other variable was observed.

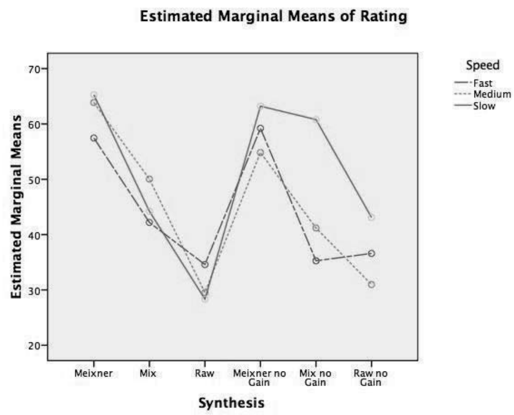


Fig 5: Main effect of speed (p<0.001) on participants ratings grouped by synthesis versions.

Figure 5 illustrates the effect of speed on synthesis version. The slow speed was ranked significantly higher for the Mix shape with no gain. The material type also affected participants' ratings of synthesis version (see figure 6). The wooden material was rated higher for "rollingness" across synthesis versions.

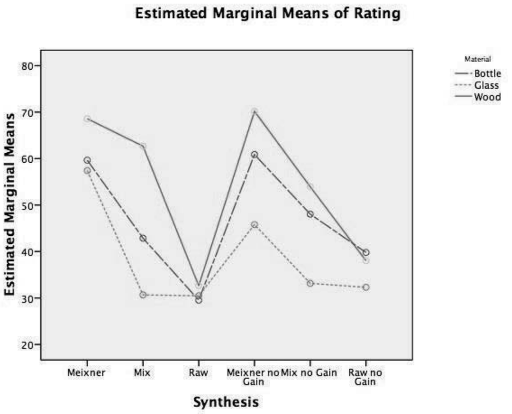


Fig 6: Main effect of material (p<0.001) on participant ratings grouped by synthesis versions.

While there was no overall effect of the gain/no gain filters on participant ratings, the interaction effects between the gain filters and the speed of the objects reveal significantly higher ratings from participants for the no gain synthesis versions at slow speeds (see figure 7).

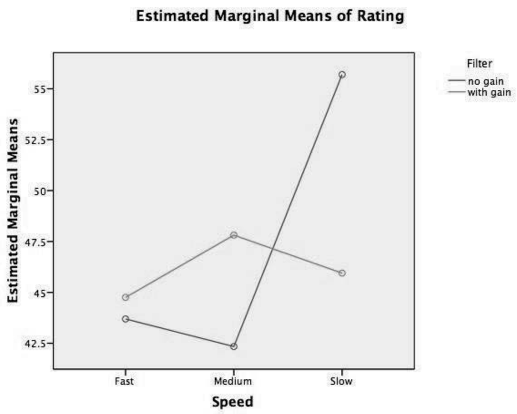


Fig 7: Interaction effect of speed * filter (p<0.001) on participants ratings grouped by rolling speed.

4.1 Qualitative Feedback

While the testing interface enforced participants to listen to each sound and provide a corresponding rating, the provision of individual comments for sounds was optional. Users tended to provide verbal feedback for individual sounds to explain a low ranking usually by commenting on distortions or unusual features perceived. For most sounds presented, participant comments were sparse, however there were a number of particular sounds that elicited verbal comments from a number of participants. For example the glass object rolling at both medium and slow speeds were described by 3 participants as "small objects dropped on a surface", "raindrops", and "sparse". This can be explained by the nature of the excitation generated by the glass object. Being very stiff and small, the glass object bounces, creating approximately periodic high amplitude impulses dominating the smaller ones due to the rolling behavior. Our synthesis algorithm only identifies the prominent impulses and discards the smaller one, which may account for the perceptual attribute as too "sparse".

The wooden object rolling at a slow speed was described as "static", "Random activity (electric activity)" "…too noise-like". At very low speeds, our algorithm fails to distinguish between dominant impacts and smaller ones, modeling them in a similar way. This generates an excitation close to being randomized, hardening the perception of a rolling object. "Static" was also a descriptor associated by 3 participants with the Raw shape.

5. Discussion of Results

It had been hypothesized that the Mix shape would be rated higher by participants in terms of realism as this synthesis technique created more accuracy in higher frequencies. Yet the shape rated significantly higher in the perceptual evaluation was in fact the Meixner shape. Furthermore post-task questionnaires revealed that users attributed higher ratings to sounds that they identified as being "lower pitched". As users were not presented with an original source for comparison their perception of the size and shape of the object was subjective and dependant on the type of object the user identified with individual sounds. Perhaps had the task been a comparison to the original recordings, the outcome may have differed. However the purpose of this perceptual evaluation was to find the best synthesis match for realism and naturalness rather than a comparative analysis to original recordings. The preference for lower frequency rolling sounds could be attributed to participants visualizing larger, heavier rolling objects, such as vehicles.

Sound stimuli with any form of distortion were attributed low rankings by users and automatically excluded from having properties of rolling. The raw shape was ranked lowest by participants and was associated with largest number of verbal descriptors for distortions. This can be attributed to the process of extracting the excitation for the Raw synthesis shape which could be improved for the next implementation.

There was no significant difference between the gain and no gain conditions for all 3 shapes (Raw, Mix, Meixner). This illustrates that, at this stage, there is no clear indication whether the gains of the modes should be modeled in the source or in the filter, which is often a design issue while dealing with source/filter models.

6. Future Work

Since there was a significant interaction effect of both speed and material it would be interesting to conduct a more detailed evaluation in these areas. Participants' preference for the no gain filter technique at slow speeds could also be an issue for further exploration. With a view to applying the synthesis rolling models to a physical task in a virtual environment, it would be interesting to evaluate participants' ability to discriminate between synthesis versions in terms of speed and material type.

As an extension of this study we intend to compare the Meixner synthesis shape, which was rated highest in this perceptual evaluation, with a revised synthesis algorithm. The interface and perceptual method will remain the same for this next auditory perception study in order to refine the synthesis technique. Adding haptic feedback will then be considered to create a more immersive virtual environment and enable exploration into both perception and action interactions for rolling across audio-haptic modalities.

References

[1] W. W. Gaver, What in the world do we hear? An ecological approach to auditory source perception. Ecological Psychology, 5(1), pp. 1-29, 1993.

[2] W. W. Gaver, How do we hear in the world? Explorations in ecological acoustics. *Ecological Psychology*, 5(4), pp. 285-313, 1993.

[3] P. R. Cook, "Physically informed sonic modeling (phism): Synthesis of percussive sounds," *Computer Music Journal*, vol. 21, no. 3, pp. 38–49, 1997.

[4] K. van den Doel, P. G. Kry, and D. K. Pai, Foleyautomatic: physically-based sound effects for interactive simulation and animation, *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pp. 537–544, 2001.

[5] J. A. Ballas, Common factors in the identification of an assortment of brief everyday sounds, *Journal of Experimental Psychology: Human Perception and Performance*, Aug Vol 19(4), pp. 250-267 1993.

[6] C. Guastavino, Categorization of environmental sounds, *Canadian Journal of Experimental Psychology*, 60(1), pp. 54-63, 2007.

[7] H. Warren, and R. R. Verbrugge, Auditory perception of breaking and bouncing events: A case study in ecological acoustics, *Journal of Experimental Psychology; Human Perception and Performance*, 10, pp. 704-712, 1984.

[8] K. van den Doel, D. K. Pai, T. Adam, L. Kortchmar and K. Pichora-Fuller, Measurements of perceptual quality of contact sound models. *Proceedings of the International Conference on Auditory Display*, Kyoto, Japan, pp. 345—349, 2002.

[9] C. Stoelinga, A psychomechanical study of rolling sound, *PhD Thesis*, Technische Universiteit Eindhoven, 2007. Available: http://pastel.paristech.org/2368

[10] M. Lagrange, G. Scavone and P. Depalle, Time-domain analysis / synthesis of the excitation signal in a source/filter model of contact sounds, *Proceedings of the International Conference on Auditory Display,* Paris, France, 2008.

[11] M. Rath, "An expressive real-time sound model of rolling," in International Conference on Digital Audio Effects, 2002, pp. 165–168.

[12] M. Rath and D. Rochesso, "Continuous sonic feedback from a rolling ball," *IEEE Multimedia*, vol. 12, pp. 60–69, 2005.

[13] R. Badeau, G. Richard, and B. David, Performance of esprit for estimating mixtures of complex exponentials modulated by polynomials, IEEE Transactions on Signal Processing, vol. 56, pp. 492–504, 2008.

[14] J. Laroche and J. L. Meillier, Multichannel excitation/filter modeling of percussive sounds with application to the piano, IEEE Transactions on Speech and Audio Processing, vol. 2, pp. 329–344, 1994. Effects Conference, 2008.

[15] B. D. Brinker, E. Schuijers, and W. Oomen, Parametric Coding for High-Quality Audio, *112th Convention of the Audio Engineering Society*, May 2002.

[16] K. van den Doel, Physically based models for liquid sounds, *ACM Transactions on Applied Perception.* 2 (4) 2005.

[17] E. Vincent, M. G. Jafari, and M. D. Plumbley, Preliminary guidelines for subjective evaluation of audio source separation algorithms, *Proceedings of the ICA Research Network International Workshop*, Liverpool, UK, 2006.

Sample, Mix and Tile: an enactive approach to graphics on objects

Daniel Saakes Walter Aprile
Delft University of Technology
E-mail: d.p.saakes@tudelft.nl, w.a.aprile@tudelft.nl

Abstract

We present an enactive approach to the creation of graphics on objects in art and design. The approach makes use of the richness in the world and consists of mixing and reusing existing visual materials from the designers' collection, both physical and virtual. By projecting the so created graphics on physical objects the interaction between graphics and shapes can be explored.

Whereas current computer support for graphics focuses on creating pixel-perfect images for presentations and manufacturing we focus on the early stage of design wherein many, rough, representations are made to explore the solution space.

Our experience shows that in this early stage designers benefit from the enactive approach to generative activities of graphics work in object design.

1. Introduction

In this paper we report on an alternative approach to generating and exploring graphics on physical objects and in physical spaces. Our approach has two aims: firstly, using the rich natural perceptual motor skills to create the graphics rather than a mediated interface like a window icon mouse pointing device (WIMP) interface. Secondly, our approach aims to make graphics experiential in their intended context: on the physical object, in the space, outside of the limits of a screen based representation.

Our application field lies in the early stages of designing objects. In these early stages designers and artists make fluent and extensive use of various types of visual materials. They browse magazines, maintain collections of samples and example products in their search for inspiration. Thoughts are expressed in sketches, moodboards, sample sheets and models that are quickly produced out of foam, clay and paper. Often designers make many sketches and models that are never shown to client. The designer's intent in creating these representations, is to explore the solution space and to support his/her internal thinking process.

The way designers work with visual media in the early stages of design is completely antithetic to how visual media are approached in mainstream computer tools such as 3D visualization tools, 2D vector drawing and bitmap editing software. These applications contain many tools to create and refine graphics and invite the user to make pixel-perfect presentations. Designers variously use these tools at latter stages of the design process with the intent of making presentations to share their designs with the larger group of stakeholders involved in the design process, or to specify designs for rapid prototyping, printing or other means of manufacturing. Typically designers make only a few of these representations and refine them during the various stages of detailing their concepts.

Computer-based graphics manipulation is usually restricted to a 2D screen and printing output. People have difficulty in linking the flat and freely scalable images to their intended application in the spatial and material world [8]. We've seen designers working on graphics to be applied on bicycles print and tape their graphics on a physical bicycle frame in order to match the graphics with the curved tubing. Ceramic artists describe the moment of opening the oven after firing their objects as magical: although they have already made many drawings and computer visualizations, that is the moment when they can experience their object for the first time in color and volume. The end-result of a product often looks completely different from what it looks like during the various stages of design.

Yet the early stage of design could benefit from the ease and speed of computer based manipulation of graphics, colors and patterns. The discrete set of visual sample materials designers often have at hand, could be extended in a way similar to mixing pigments on a painters' palette. Moreover, projected augmented reality could provide a means to explore graphics in the context of physical shapes and environments.

In this paper we report our experiences with a tool called Skin. Skin projects computer generated graphics on physical objects and thereby will alter their appearance.

Earlier [11] we reported on the development of the tool and why exploring materials in context of the product shape is important, in this paper we will focus on the mechanics of the user interaction in the creation of graphics.

First we discuss related work. Then we present the design of Skin, the user interaction mechanics and its configurability. Then we discuss our experiences using Skin.

2. Related work

Many efforts to adapt computer tools to support the early stages of design focus on shape, namely on 3D modeling or sketching. [e.g. 3]. Another research line involves painting and handling visual materials in a physical way that transcends screen-based interaction.

The I/O brush [10] is a drawing tool for children to explore colors, textures and movements found in everyday materials by picking them up and drawing them with them on a digital canvas. The physical brush eliminates the need for a predefined digital palette. The I/O brush encourages children to explore their surroundings in the process of creating their drawings.

Hoeben's digital sketchbook [5] is a tablet PC application that aims to incorporate the qualities and advantages of physical sketchbooks with regards to handling large amount of sketches. His sketchbook contains no traditional UI widgets. The elements of the interface are related to the designer's own preferences and to the cultural context of sketchbooks: for example, each page contains a small "inktest" that is used to select a drawing style. This inktest area is drawn by the designer in his/her own personal style and so introduces far less distraction than traditional user interface.

Cabinet [6] merges the physical and digital collections of materials that many designers like to keep. The large tablet interface supports users in organizing materials in spatial layouts rather than in a linguistic hierarchical structure. The interface is designed around the designers' way of structuring thoughts. A built-in digital camera digitizes physical materials on the fly to bridge the separation of physical and virtual media. Cabinet makes use of a larger scale than a 2D screen interface, which allows groups of people to interact with the system.

The tools we have just mentioned support the early stages of the design process and help the designer to manage the visual richness of the everyday world. They show that the results generated with the tools do not have to be a single final document.

A few tools in the field of spatial augmented reality [1] aim to support designers by augmenting physical models with computer generated light [7, 8, 13]. The texture brush of the Atelier project [7], for example is a system that allows architects to draw textures on scale models of buildings with a brush. However, these systems are about creating presentations. They make use of 3D visualization software or traditional WIMP interfaces to select and generate graphics.

Piper's illuminated clay [8] is one of the few systems that allow designers to modify shapes, in his case a clay model of the landscape and observing the implications of a change in place on the model.

3. Designing Skin

3.1 Goals and approach

Skin's design considerations are summarized in table 1. Skin aims to make use of the designer's existing collection of materials, to mix and modify them rather than to create new graphics. The user's experience and the system interaction should be more like sketching than using CAD.

participative	All stakeholders can participate and explore the solution space first hand, not mediated
thinking, sharing	A tool that supports the designer's though process and allows sharing thoughts within the design team
explorative	Generative, fast, non committal, similar to sketching.
existing materials	Makes use of designer's own materials (models, drawings, samples)
natural skills	Makes use of the designer's natural skills of manipulating materials.
context	Show the interactions between shape and material

Table 1 Skin's design considerations.

In this early stage of a design project, many stakeholders are involved and they do not necessarily have a design background: users, clients, marketing people, engineers... all have to share in the design.

Through contextual inquiry we have found that these people want/need to be involved in the design of graphics but do not necessarily have the means to express themselves or ways to explore graphics by doing. Sometimes, for instance, the designers' clients supply pictures of a particular style they want to achieve in the product. Skin aims to make these participants active contributors to the design process.

3.2 Skin

figure 1. The Skin setup consists of two interaction spaces. To the right and front you see the material palette whose function is acquiring and mixing visual materials. A webcam mounted on a swiveling arm grabs the materials. The object table on the left and back contains a projector that illuminates objects on the table with colored light.

Skin consists of two components as shown in figure 1: an interaction space to mix materials: “the material palette” and an interaction space for physical models, the “object table”. Both spaces are placed near to each other, so users can interact with both concurrently or switch easily between them.

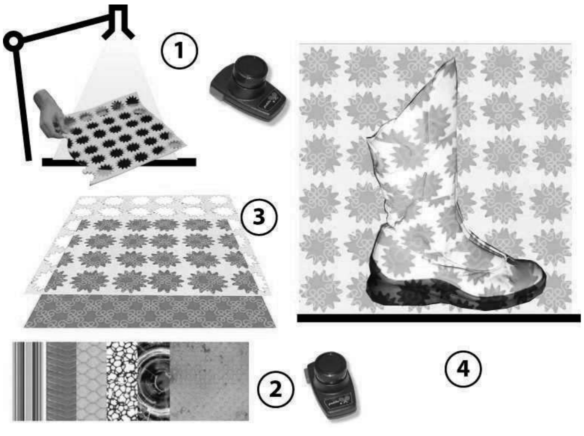


figure 2. An overview of the Skin graphics pipeline: the material table and the paddle controller to scale and tile the materials (1). A digital collection of materials and the associated paddle controller (2). The physical materials are superimposed over the digital materials (3) and projected on the object table (4).

The material palette consists of small table with a camera mounted on a swivel. The camera captures materials on the table. This table is large enough to hold a small amount of materials, as shown in figure 3, among them cut-out graphics and samples from manufactures. The material palette has a paddle controller to adjust the scale and automatically tile the projected graphics.

figure 3. Mixing and tiling of artwork. In this example black will be made translucent through chroma keying.

Optionally a second paddle controller controls a digital collection of visual materials. The graphics are stored on a USB stick and a button on this paddle controller flips through the materials like a slideshow. Rotating the paddle controller adjusts the scale of the graphics. Then, the physical graphics build on the material palette are superimposed on the digital graphics by means of chroma keying.

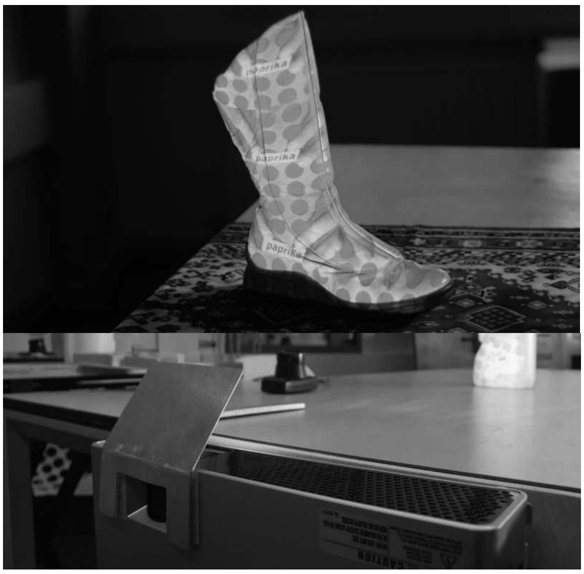


figure 4. The object table. A projector projects the created graphics through a mirror over the table in such a way that the objects on the table are illuminated while the table surface is not.

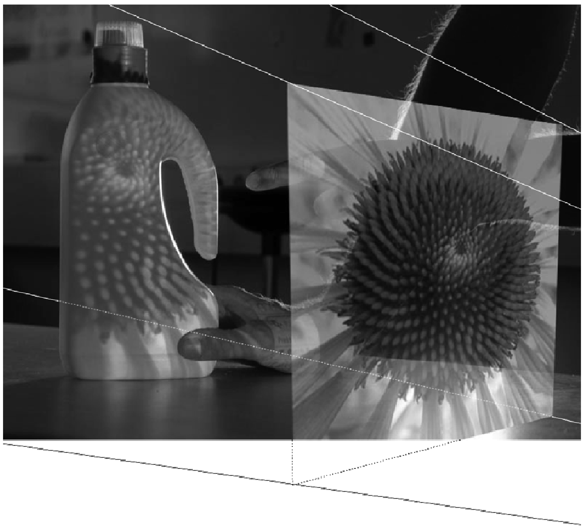


figure 5. Moving an object over the table to align shape features to detail in the graphics. The display frustum originates on the right (not shown).

The second component of Skin is an interaction space for physical models. We call it “the object table”. The object table is a common table under which a video projector is mounted vertically. A small mirror bends the optical path so that the projection frustum skims the table surface and illuminates only the objects on table.

Differently from typical augmented reality applications, Skin has no knowledge of the objects on the table, so any object can be used, as long as it is light-colored enough to function as a projection surface.

3.3 Implementation

The current prototype employs an off the shelf LCD projector and a webcam. The software runs on an Apple Macintosh OSX machine and is written in the free available Quartz Composer program [9]. Quartz Composer is intended for writing screensavers and video effects. It manipulates images and video efficiently through pixel and vertex shaders on the graphics accelerator.

The paddle controllers are originally parts of the classic Atari 2600 computer system. The potentiometer in each controller has been replaced by a Phidget [4] rotary encoder in order to make it turn endlessly.

3.4 Creating graphics with Skin.

The light from the projector forms a pyramid shaped frustum originating at the edge of the table and extending over it, as depicted in figure 5. To translate graphics on the object, the object is moved perpendicular to the axis of the frustum. Likewise rotating the object on this plane causes a rotation. Zooming graphics is achieved to a certain extent by moving the object towards or away from the apex.

Obviously, projecting without calibration on non planar objects will warp the graphics over the object’s surface. Traditionally, this is treated as an imperfection. However, viewers can easily see through this, just as they can deal with vagueness and ambiguity in sketches. Moreover, we see this as a feature that will provide serendipity and insights in the shape/graphics interactions.

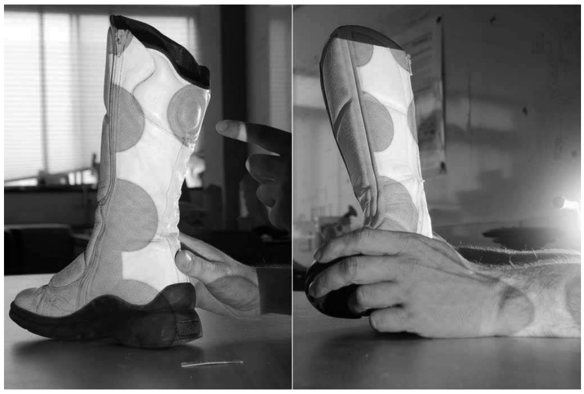


figure 6. By holding the boot in a certain way, the graphics nicely line up with the shape features, on the left with a circle pattern and on the right in the centre symmetry axis.

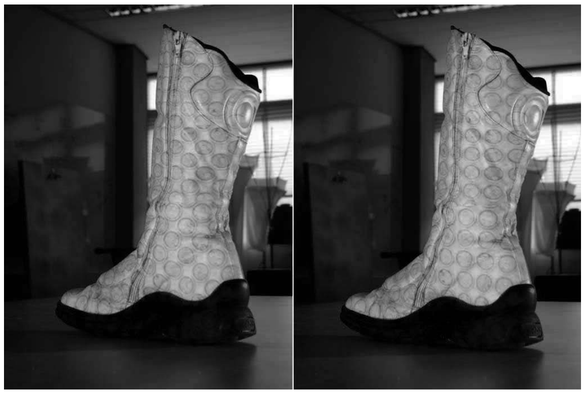


figure 7. Rotating the circle pattern on the material palette (see figure 3.) and experiencing the result on the object.

The physical materials are cropped by the viewing frustum of the camera and can furthermore be altered by changing the height, position and angle of the camera.

The user is encouraged to manipulate the object over the table and alter the complete visual effect.

Graphics can be manipulated by translating and rotating the source objects below the camera to make them fit on the object. A physical material can be made seamless tile-able for instance, by adjusting the position and rotation, and the height of the camera.

4. Experiences

We have applied Skin in practice in various design situations, working with individual artists, teams of designers, marketing managers and users. For each application we contextualized the Skin configuration to the users' need. For instance, in creative group workshops in packaging we have worked with digital images as a non-scalable background. The focus was on the interaction with physical vegetables and cut-out branding graphics in the design of potato chips packages. On the contrary, when working with artists that maintained large collections of digital materials, physical sources of patterns have been eschewed.

The creation of graphics with physical materials is intuitive and users immediately start experimenting. It invites users to play with fabrics, to actively search their surroundings for materials and to make collages from graphics found in magazines.
Generally people start playing by placing graphics on the material table and experience the effects on the objects on the table. They usually flip quickly through a lot of patterns and play with the scale button. We consider this behavior "browsing" and explorative creativity. There seem to be a few specific scales on which the graphics interact with the details on a particular shape.
A second type of behavior emerged during packaging workshops, namely the creation of collages from layered cut-out graphics. For instance a background of a meadow combined with graphics of flowers and brand-specific graphics. We consider this behavior to be goal orientated creativity.

As we have remarked before, projecting graphics in this way on non planar objects will warp the graphics over the object's surface. That the visualization as a whole is not perfect due to the warping of the images, seem not to matter: people tend to zoom in on specific areas they like.
In a workshop on ceramic design a traditional decoration warped over a bowl created a surprising effect when the bowl was moved through the projection. This effect sparked a new series of oddly placed and warped decorations.
Often new ideas are found by accident. For instance when projecting graphics on a domestic appliances the power plug and wire were illuminated with graphics, and that sparked new ideas of textures on wires and plugs. These accidents are less likely to occur when using the traditional CAD tools.

However, for novel users, the concept of having a digital collection as well as a physical means of creating graphics is less understandable. Compared to the open and physical interaction with both the models and materials makes the digital collection remains a hidden and less accessible feature.

4.1 Exploring shapes

figure 8. A few simple shapes built with the surface-toolkit. On the right a larger mouth is added to an existing bottle.

In a workshop that aimed to design a tableware we've seen designers buy a great amount of white ceramic objects and play with them. This was an effective way to change and modify shapes in order to see the effect of a particular graphic on various curved surfaces.
In other workshops we've tried to integrate Skin with paper and clay prototyping. There we discovered that the shape prototyping invites people to make beautiful and detailed prototypes instead of exploring shapes roughly.
Some creativity methods such as interaction relabeling [2] take a fixed, inspirational set of shapes to explore interaction by making a forced metaphor to the design problem. Sanders developed a foam modeling toolkit [12] to have users design their ideal product, and while building they where better able to express their wishes. The limited form-freedom of these toolkits support participants to think and explore rather than to focus on detailing.
Likewise, for our purposes we might need a toolkit that allowed the easy creation of surfaces. We have started to experiment with a toolkit of thin white plastic sheets , see figure 8, that can be bent and put together with Velcro.

4.2 Documenting design

Graphics on Skin exists at the moment it is created. In order to get back to an earlier design the user needs to recreate the graphics in the material space and to place the objects on the table.
At the moment we support no capturing for documentation in Skin, we simply ask users to take photos. However, in our experience we often find that people forget to take pictures. Skin has need for an simple way to capture a design, that will not break the users' creativity flow.

5. Future work

Our approach is sketchy and presents a limited feature set. We think it is great for rough and early stage prototyping, but in its current form less suitable for later stages of design.

Skin has no knowledge of the objects on the table and colors everything with the same graphic pattern.
Often objects do not consist of a single material but of a few. Multiple materials are supported to a limited extent by arranging multiple materials on the material space, or cutting masks from paper. A solution to distinguish areas on the models would extend the area of application.
A way to store created graphics in a memory bank would facilitate expert users; they could build up graphics in layers similar to how audio samples are layered in multitrack sound editors. Likewise a similar feature would allow users to quickly go back to earlier created graphics.

6. Discussion

With the Skin prototype we uncovered a few ways in which tools aimed at the early stage of object design might support designers in their search for colors and graphics. The roughness of the tool and the way of handling materials enhances explorative creativity and triggers creative accidents that are less likely with traditional computer tools.

We make use of the designer's existing skills and the richness of the world, and we add a little computer support, just were it is needed. Our experiences with Skin might point to future graphic tools that aim to support designers in the early stage of design.

7. Acknowledgements

The authors like to thank Aernout Peeters for his photos that have been used in this paper and Aadjan van der Helm for his suggestions.

References

[1] Bimber, O. Raskar, R., Spatial Augmented Reality. Merging Real and Virtual Worlds. A.K. Peters Wellesley, Massachusetts. 2005

[2] Djajadiningrat, J.P., Gaver, W.W., Frens, J.W. Interaction relabeling and Extreme Characters Methods for Exploring Aesthetic Interactions. *DIS '00: Proceedings of the 3rd conference on Designing interactive systems.* 66-71, 2000

[3] Do, E. Y-L Drawing marks, acts, and reacts: Toward a computational sketching interface for architectural design. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*. 149-171. 2002

[4] Greenberg, S., Fidchett, C. Phidgets: Easy Development of Physical Interfaces through Physical Widgets.*UIST.* 2001

[5] Hoeben A., Stappers P.J., Direct talkback in computer supported tools for the conceptual stage of design, *Knowledge based systems*, 18 (2005) 407-413, 2005

[6] Keller, A.I., Hoeben, A., van der Helm, A. Cabinet: Merging designers' digital and physical collections of visual materials. *Personal and Ubiquitous Computing.* 2005

[7] Matkovic, K. (2005). Dynamic Texturing of Real Objects in an Augmented Reality System proceedings of IEEE Virtual Reality Conference

[8] Piper, B., Ratti, C., Ishi, H. Illuminating Clay: A 3-D Tangible Interface for Landscape Analysis". *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems.* 355-362. 2002

[9] Quartz Composer. Apple Computers. http://developer.apple.com/graphicsimaging/quartz/quartzcomposer.html retrieved July 2008

[10] Ryokai, K., Marti, S., and Ishii, H. I/O brush: drawing with everyday objects as ink. *CHI'04: Proceedings of the SIGCHI.* ACM Press, New York, NY, 2004

[11] Saakes, D.P. (2006). Material light: exploring expressive materials. *Personal and Ubiquitous Computing*. 10(2). 144-147.

[12] Sanders, E.B.-N., William C.T. Harnessing People's Creativity: Ideation and Expression through Visual Communication. In *Focus Groups: Supporting Effective Product Development.* Langford J and McDonagh-Philp D (Eds.) Taylor and Francis. 2001

[13] Verlinden, J.C., de Smit, A., Peeters, A.W.J., van Gelderen, M.H. (2003). Development of a Flexible Augmented Prototyping System. *Journal of WSCG*, 11(3). 496-503. 2000.

Virtual Reality Simulator of Transurethral Resection of the Prostate

Padilla Miguel A.*, Teodoro Sergio*, Soriano David*, Reyes Bartolome*, Altamirano Felipe*, Durán Sergio†, Arámbula Fernando*

(*) Image Analysis and Visualization Group, CCADET, Universidad Nacional Autonoma de Mexico, P.O.Box 70-186, Mexico, D.F., 04510

(†) Instituto Nacional de Rehabilitación, INR, México

E-mail: miguel.padilla@ccadet.unam.mx,fernando.arambula@ccadet.unam.mx

Abstract

In this work we present the current state of development of a virtual reality surgery simulation system for training of Transurethral Resection of the Prostate (TURP). The prototipe consists on a virtual environment of part of the urinary system which includes 3D models of the bladder, the prostate and the pennile urethra. The system simulates in real-time the endoscopic view of the resectoscope, allows realistic virtual navigation and exploration inside the anatomic structures of the virtual patient. The simulator includes a deformable tissue model of the prostate for simulating tissue resection and deformation. The interaction between the trainee and the system is done with a mechatronic interface that emulates a real resectosope and allows to perform the most important movements of the surgical tool during a TURP. We are currently working in incorporating tactile feedback to the interface by adapting a haptic device to the mechanism.

1 Introduction

The prostate gland is located next to the bladder in human males, with the urethra running from the bladder neck through the prostate to the penile urethra (Figure 1). A frequent condition in men above 50 years old is the benign enlargement of the prostate known as Benign Prostatic Hyperplasia (BHP), which in some cases results in significant blockage of the urinary flow. The standard surgical procedure to treat a hypertrophied prostate gland is a minimally invasive surgical procedure called Transurethral Resection of the Prostate (TURP). It essentially consists of the removal of the inner lobes of the prostate in order to relieve urinary outflow obstruction. Mastering the Transurethral Resection of the Prostate (TURP) technique requires a highly developed hand-eye coordination which enables the surgeon to orientate inside the prostate, using only the monocular view of the lens of the resectoscope (Figure 2). Currently TURP is taught through example from an experienced surgeon. The resident of urology has very restricted opportunity to practice the procedure.

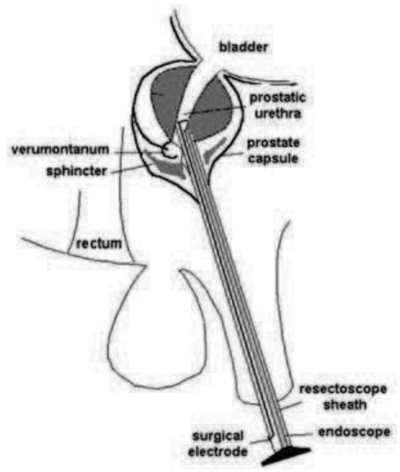


Figure 1: Schematic view of the prostate and resectoscope interaction during TURP.

An alternative that is currently gaining acceptance for training several minimally invasive surgical procedures is the use of computer surgery simulators. Surgery simulators have shown to be usefull to help novice surgeons to acquire the skills in shorter periods of time, with less risk for the patiens, and in an economically suitable manner [18].

Several interesting works have been reported in the past, for example, Gladilin et. al. [4] reported a biomechanical model, based on FEM, of soft tissue deformation for surgical planning and prediction of human facial shape after craniofacial surgery, where the results is crucial for the patients. Szkely et. al. [17] described the development of a special parallel hardware implementation of finite element algorithms for tissue deformation simulation, and developed a prototype for a laparoscopic gynaecology simulator. Khnapfel et al. [6] used both, FEM and mass-spring models, on visually realistic endoscopic surgery training systems. They use FEM or mass-spring model, depending on the scenario and the situation. Brown et al. [2] reported a system for a microsurgery simulator which performs small blood vessel suturing. Radetzky and Rudolph [14] described ROBO-SIM a software for simulating tumour removal in neurosurgey; DiMaio and Salcudean [3] presented a technique for simulating needle insertion in soft tissues for percutaneous procedures.

Nowadays, exist some commercial systems, for example the Immersion company [8] commercialize simulators for training endoscopic, endovascular, laparoscopic and arthroscopic procedures. The Simbionix company [15] commercialized systems for simulating interventional endovascular procedures for cardiology, gastrointestinal endoscopic procedures, laparoscopic surgical procedures, cystoscopy and ureteroscopy procedures in urology, among others.

In the scope of urology, Manyak et al. [7] reported the construction of a virtual reality surgery simulation system of the lower urinary track. They considered only the surface of the urinary track, reconstructed from the Visible Human Project dataset [9] with texture mapping for visual realism. However, the prostatic urethra behaviour depends on the conditions of the tissue from the capsule to the urethra. As a consequence, a volumetric model of the prostate should be considered in order to simulate realistic TURP procedures. Sweet, et al. [16] reported the experience of using the TURP virtual training system described in [10]. They studied the effectiveness of translating the skills acquired in their virtual environment to the operating room. The surgery simulation system for TURP reported in [10] uses an image-based approach for simulating bleeding when the resecting loop contacts surface vessels. The loop triggers precalculated movies of blood flow, that is then oriented and mapped on the virtual environment. However, how they model tissue deformation is not clear; moreover as in [7] they used only the surface representation of the urethra.

In [1, 12] we reported the development a 3D deformable volumetric model of the prostate for TURP simulation that involves tissue deformation and resection, considering the gland as a viscoelastic solid. In this work we describe the development of the virtual environment which includes the full virtual TURP scenario and physical interaction between the simulator and the trainee. Section 2 of this paper describes the design of the virtual resectoscope interface; section 3 describes the development of the virtual environment, the interaction scheme and tissue modelling; finally in section 4 we present the conclusions and future perspectives of this work.

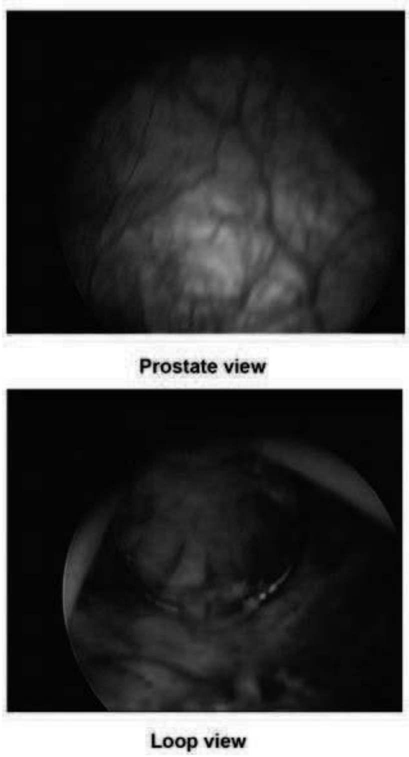


Figure 2: Endoscopic view of the prostateand the resecting loop of the resectoscope.

2 The resectoscope device

In order to obtain a realistic simulation of the most important movements of the surgeon during a TURP, a mechanism was designed based on a disk-ring array [11]. We decided to reproduce only the five most important degrees of freedom of the real movements during a TURP (Figure 3). We ignore two additional translational movements of the resectoscope sheath. In opinion of an expert urologist, the five movements chosen are enough for having a realistic reproduction of the real movements during cistoscopies and TURP procedures.

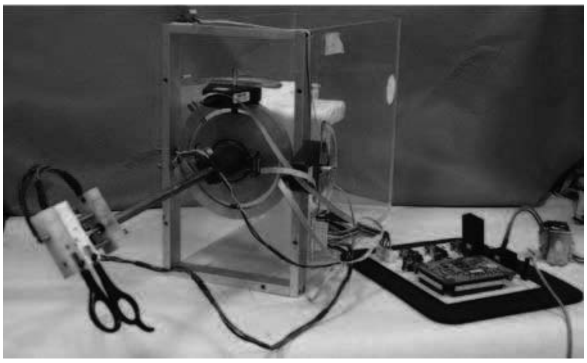


Figure 3: View of the mechatronic device that emulates a real resectoscope.

2.1 Signals sensing

Three of these axes are rotational and the other two are linear displacements of the resectoscope (Figure 4). Optical encoders are used in order to sense each of the three rotational movements; the linear displacement of the surgical tool is measured with a linear precision potentiometer; the resecting loop movement is measured with a two linear Hall effect sensors array and two permanent magnets.

Linear displacement of the sheath of the resectoscope is measured with a linear precision multi-turn potentiometer. The output voltage of the potentiometer varies according to the position of the resectoscope from 0 to 23 cm, which corresponds to the useful resectoscope displacement range. We programmed in a microcontroller a linear interpolation in the range of [0-23] cm in order get real-time displacements of the sheath.

The resecting loop has a linear movement in the range o [0-36] mm. This distance is measured with an array of two linear Hall effect sensors and two permanent magnets. We precalculated and stored in another microcontroller a lookup table that models the observed nonlinear signal of the Hall effect sensors array, where the lookup table consist on 148 values that corresponds to a resolution of 0.5mm of the nonlinear signal. Finner displacements behind 0.5mm are calculated in real-time by doing linear interpolation over the precalculated values. Optical encoders from USDigital (HEDS-9040 and HEDS-9140 models) [19] are used in order to sense each of the three rotational movements. Each encoder is placed in each rotation axe, so with this arrangement we can measure the direction and the angle rotated by the user on each axis. One dedicated microcontroller is attached to each optical encoder in order to sense in parallel all rotations in real-time.

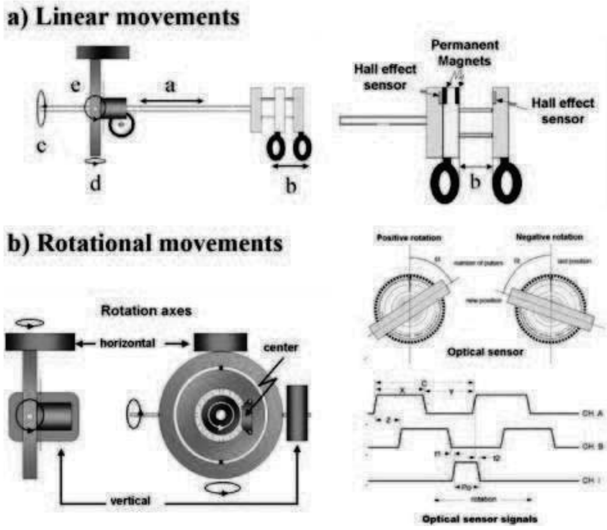


Figure 4: Hall effect sensors array for controling the resecting loop.

2.2 Embebed system for signals monitoring

Monitoring of the movements is performed with an embedded electronic system consisting on five microcontrollers: 1) Four PIC microcontrollers ([5])for digital signal monitoring of the optical sensors for rotations, three for sensing and one for collecting rotational data; 2) one microcontroller (Rabbit 3000 8 bit microcontroller at 7,4 MHz with analog-digital converters [20]) for the analog signals of the linear potentiometer and hall-effect sensors, for translations; 3) a LP3500 single-board computer for collecting all data and send it to the graphical workstation [20].

The five microcontrollers in the embedded system run together in parallel in order to monitor the movements in real time; the system sends the movement information in the form of commands to the virtual model. The real-time resectoscope movements consequently reflect the interaction between the surgeon and the tissue model.

3 The RTUP Simulation

3.1 The Virtual Environment Modelling

We modeled the anatomical structures of the urinary system involved in TURP procedures: the bladder, the prostate and the pennile urethra in order to simulate virtual explorations (cistoscopies) and resections. The prostate mesh was reconstructed from a computer set of segmented transuretral ultrasound images of a real patient. Since computing tissue behaviour for the prostate

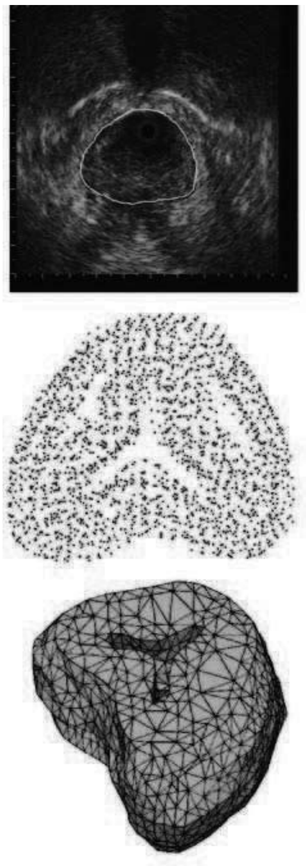


Figure 5: Adaptive model of the prostate volume. From top to bottom: a) Segmented ultrasound image. b) Morphological shape interpolation and adaptive sampling. c) Adaptive tetrahedral mesh.

is a demanding task that depends on the complexity of the mesh, we developed an adaptive reconstruction algorithm that modulates the resolution of the Delaunay tetrahedrization with respect to the 3D distance field of the external shape previously interpolated from the original stack of segmented contours (Figure 5), in order to have a visually realistic model but at the lowest possible mesh complexity for numerical computing. The pennile urethra and the virtual resectoscope consist on 3D surface meshes built with a 3D modeling software. The bladder consists also on a 3D triangular mesh reconstructed from segmented images of the Visible Human Dataset (Figure 6).

We also modeled some other aspects of a real TURP with enough visual realism such as: the endoscopic view from the resectocope, the ilumination of the resectoscope lamp and the tissue texture appereance by using procedural textures techniques, and the urethra sphincter (Figure 7).

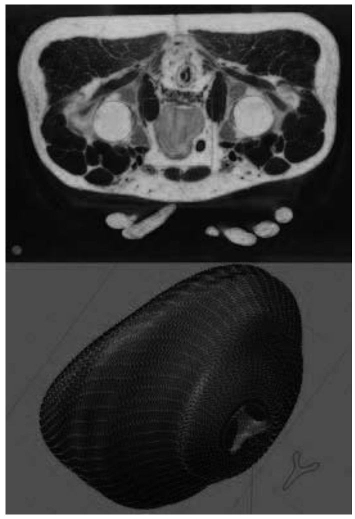


Figure 6: Model of the bladder. a) A VH cryosection with the bladder identified for manual segmentation. b) 3D bladder mesh obtained with triangular reconstruction.

3.2 Collision Detection for Interaction

We developed a collision detection algorithm based on the representation of triangular meshes with hierarchical trees of bounding spheres. In order to speed up collision detection computations, our implementation allows an interruptible scheme with a predifined maximum time for computing collisions.

After a collision is detected the soft tissue must slightly deform before the tissue resection occurs. Tissue deformations result from the reacting forces produced after the collision. To calculate the external reacting forces we used a penalty-based method, where reacting forces are the forces needed to separate the penetrating objects. These forces depend linearly on the penetration depth field and the stiffness and damping properties of the soft model (Figure 8). For calculating the penetration field we use the history of the last collision stored: we obtained the signed distance field of the colliding submesh of the resectoscope with respect to the submesh of the prostate in order to use the distance as measure of penetration; the normals of the prostate submesh define the direction of the reacting forces; the sign is used to discriminate the triangles of the resectoscope in contact that not penetrate the prostate from the penetrating ones. More details of the collision detection mechanism are available in [13].

Figure 7: Virtual environment of TURP with endoscopic view from the head of the Resectoscope. From top to bottom: a) View of the pennile urethra. b) View at the prostate entrance near the sphicter. c) Inside view of the prostate. d) Inside view of the bladder.

3.3 Tissue Modelling

Due to its simplicity we used a well known mass-spring method for modeling the deformable behaviour of the soft tissue of the prostate, where the dynamic system is determined by the following equation of motion:

$$m_i \frac{d^2\mathbf{x_i}}{dt^2} + \gamma_i \frac{d\mathbf{x_i}}{dt} + \mathbf{g_i}(t, x_i) = \mathbf{f_i}(t, \mathbf{x_i}) \qquad (1)$$

where m_i is the mass of the node N_i, at cartesian coordinates x_i; γ_i is the damping coeficient of the node; g_i is the internal elastic force over N_i; f_i are the external loads acting over N_i. The internal elastic force acting over N_i is:

$$\mathbf{g_i} = \sum_{j \in N(N_i)} \mu_{i,j} \frac{(\|\mathbf{x_i} - \mathbf{x_j}\| - {l_{i,j}}^0)(\mathbf{x_i} - \mathbf{x_j})}{\|\mathbf{x_i} - \mathbf{x_j}\|} \qquad (2)$$

Where $\mu_{i,j}$ is the stiffness of the spring that connects N_i and N_j ($\forall N_i \in N$, where N are the neighbours of N_i); ${l_{i,j}}^0$ is the initial length of the spring. Deformations result from the internal elastic energy produced by the spring arrangement over the tetrahedral mesh and the external forces applied on the prostate surface.

Since computing tissue deformation is another crucially demanding task we implemented the equation system solving with a texture shader on a GPU (Graphical Processing Unit) by using Cg language. Changing the notation of equation 1 in a matrix state form, we programmed a simple Euler solver, defined as the following set of state equations:

$$\mathbf{M\ddot{X}_t} = \mathbf{F_t} - \mathbf{D\dot{X}_t} - \mathbf{G_t}, \qquad (3)$$

$$\mathbf{\dot{X}_{t+\Delta t}} = \mathbf{\dot{X}_t} + \mathbf{\Delta t \ddot{X}_t}, \qquad (4)$$

$$\mathbf{X_{t+\Delta t}} = \mathbf{X_t} + \mathbf{\Delta t \dot{X}_{t+\Delta t}} \qquad (5)$$

where $\mathbf{M}$ is the mass matrix of the system, $\mathbf{F}$ represents the loads vector, $\mathbf{D}$ is the damping matrix, and $\mathbf{G}$ is the internal elastic force vector.

By transforming the system to a state matrix representation, matrices $\mathbf{X}$, $\mathbf{M}$ and $\mathbf{D}$ are loaded to the shader as input textures; for computing the inner elastic forces $\mathbf{G}$ the connective texture $\mathbf{N}$ is also passed to the shader, where $\mathbf{N}$ contains the pointers to the neighbours of each node. Then, using the connective information and using a render to texture technique, the computations of the new nodes positions are stored in an extra framebuffer (other than normal rendering buffer) and send it back from the GPU to the CPU at each time step. In this manner, the CPU and the GPU work in parallel, in order to achieve real-time interactions. CPU-GPU parallel computing is sincronized with a multithreading scheme. For tissue resection we are currently developing a mechanism for local and real-time mesh refinement near the contact zone in order to reduce the decimation effect after tissue cutting. Tissue cutting is performed as two operations: removing geometrical elements within the cutting zone of the resecting loop and adding the inner triangles exposed by the cut.

4 Conclusion and Future Work

In this communication we reported the recent advances in the development of a computer simulator for training of RTU. Currently the system includes a real full scenario of TURP and passive real-time interaction with a mechatronic interface similar to a real resectoscope, without force feedback.

Thanks to the GPU-CPU parallel multithreading computing scheme and the adaptive reconstruction method of the prostate, the simulator is able to run

Figure 8: Tissue deformations due to interaction with the resectoscope.

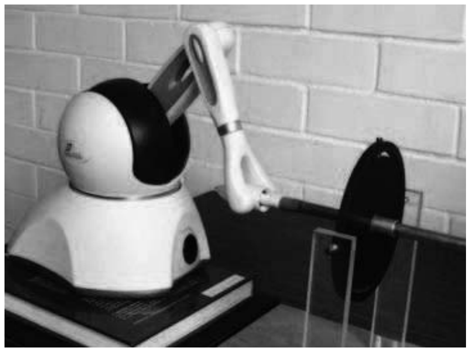


Figure 9: Scheme for adapting the resectoscope interface with an haptic robot for tactile feedback.

within a frame rate close to 30 Hz, which is quite acceptable for real-time interactions.

We are currently working in enhancing the tissue resection mechanism and in animation techniques for adding to the user the possibility of practicing tissue coagulation.

We are also working in the design, development and instrumentation of a more precise and realistic mechatronic interface which will include the instrumentation of the pedals for controlling coagulation and resection. The new resectoscope interface will be adapted to work in conjuntion with a PHANTOM (haptic robotic arm), in order to provide tactile feedback, as can be shown in the maniquin of (Figure 9). For the moment, the current integration with the maniquin shows an acceptable haptic response rate of 650 Hz for the force rendering; however, we expect that the new mecathronic interface with faster and full digital encoders will reach a force rendering rate around 800 Hz. It is also important to notice that the haptic system must be calibrated in order to have correspondance between the coordinate systems of the PHANTOM and the mechatronic device; of course, force intensity calibration with the help of the surgeon must also be done in order to have suitable tactile sensation.

We expect in the short term to have a prototype version with resection and coagulation capabilities with visual, tactile and sound feedback, and in the mid term we expect to add bleeding simulation.

References

[1] F. Arambula-Cosio, M. Padilla-Castañeda, and P. Sevilla-Martinez. Computer assisted prostate surgery training. *International Journal of Humanoid Robotics*, 3(4):485–498, 2006.

[2] J. Brown, S. Sorkin, J. Latombe, K. Montgomery, and M. Stephanides. Algorithmic tools for real-time microsurgery simulation. *Medical Image Analysis*, 6(3):289–300, 2002.

[3] S. DiMaio and S. Salcudean. Needle insertion modelling for the interactive simulation of the percutaneous procedures. In *Medical Image Computing and Computer Assisted Intervention (MICCAI), Lecture Notes in Computer Science*, pages 253–260, 2002.

[4] E. Gladilin, S. Zachow, P. Deuflhard, and H. Hege. A biomechanical model for soft tissue simulation in cranofacial surgery. In *International Workshop on Medical Imaging and Augmented Reality*, pages 137–141, 2001.

[5] M. T. Inc. *PIC16F87XA Data Sheet 28/40/44-Pin Enhanced Flash Microcontrollers.* http://ww1.microchip.com.

[6] U. Kuhnapfel, C. H.K., and H. MaaB. Endoscopic surgery training using virtual reality and deformable tissue simulation. *Computers and Graphics*, 24:671–682, 2000.

[7] M. Manyak, K. Santangelo, H. J., R. Kaufman, T. Carleton, X. Hua, and W. R.J. Virtual reality surgical simulation for lower urinary tract endoscopy and procedures. *Journal of Endourology*, 16(3):185–190, 2002.

[8] I. Medical. *http://www.immersion.com/medical/*.

[9] NLM. *The Visible Human Project*. National Institutes of Health.

[10] P. Oppenheimer, A. Gupta, S. Weghorst, R. Sweet, and J. Porter. The representation of blood flow in endourologic surgical simulations. *Stud Health Technol Inform*, 81:365–371, 2001.

[11] M. Padilla, F. Altamirano, and F. Arámbula. Virtual resectoscope interface for a surgery simulation training system of the prostate. In *3rd Workshop in Virtual Reality Interaction and Physical Simulation VRIPHYS'06*, pages 101–108, 2006.

[12] M. Padilla-Castańeda and F. Arambula-Cosio. Deformable model of the prostate for turp surgery simulation. *Computers and Graphics*, 28:767–777, 2004.

[13] M. Padilla-Castañeda and F. Arámbula-Cosio. Improved collision detection algorithm for soft tissue deformable models. In *Sixth Mexican International Conference on Computer Science, IEEE*, pages 41–47, 2005.

[14] A. Radetzky and M. Rudolph. Simulating tumour removal in neurosurgery. *International Journal of Medical Informatics*, 64:461–472, 2001.

[15] Simbionix. *http://www.simbionix.com*.

[16] R. Sweet, T. Kowalewski, P. Oppenheimer, S. Weghorst, and R. Satava. Face content and construct validity of the university of washington virtual reality transurethral prostate resection trainer. *Journal of Urology*, 172(5):1953–1957, 2004.

[17] G. Székely, C. Brechbühler, R. Hutter, A. Rhomberg, I. N., and P. Schmid. Modelling of soft tissue deformation for laparoscopic surgery simulation. *Medical Image Analysis*, 4(1):57–66, 2000.

[18] M.-D. Tsai, M.-S. Hsieh, and S.-B. Jou. Virtual reality orthopedic surgery simulator. *Computers in Biology and Medicine*, 31:333–351, 2001.

[19] USDigital. *Three Channel Optical, Incremental Encoder Modules HEDS-9040 & HEDS-9140*. Hewlett Packard; Technical Data 5965-5870E.

[20] ZWorld. *http://www.zworld.com*.

Author Index